Remarks on the necessity and implications of state-dependence in the black hole interior

Papadodimas, Kyriakos; Raju, Suvrat

Published in:
Physical Review D

DOI:
10.1103/PhysRevD.93.084049

Publication date:
2016

PHYSICAL REVIEW D **93,** 084049 (2016)

Remarks on the necessity and implications of state-dependence in the black hole interior

Kyriakos Papadodimas[1,2,*] and Suvrat Raju[3,4,†]

[1]*Theory Group, Physics Department, CERN, CH-1211 Geneva 23, Switzerland*
[2]*Centre for Theoretical Physics, University of Groningen, Nijenborgh 4 9747 AG, The Netherlands*
[3]*International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, IISc Campus, Bengaluru 560012, India*
[4]*Center for Mathematical Sciences and Applications, Harvard University, 1 Oxford Street, Cambridge, Masachusetts 02138, USA*

(Received 9 May 2015; published 28 April 2016)

We revisit the "state-dependence" of the map that we proposed recently between bulk operators in the interior of a large anti–de Sitter black hole and operators in the boundary CFT. By refining recent versions of the information paradox, we show that this feature is necessary for the CFT to successfully describe local physics behind the horizon—not only for single-sided black holes but even in the eternal black hole. We show that state-dependence is invisible to an infalling observer who cannot differentiate these operators from those of ordinary quantum effective field theory. Therefore the infalling observer does not observe any violations of quantum mechanics. We successfully resolve a large class of potential ambiguities in our construction. We analyze states where the CFT is entangled with another system and show that the ER = EPR conjecture emerges from our construction in a natural and precise form. We comment on the possible semiclassical origins of state-dependence.

DOI: 10.1103/PhysRevD.93.084049

I. INTRODUCTION

Recent work by Mathur [1], Almheiri *et al.* [2,3] and then by Marolf and Polchinski [4] has sharpened the information paradox [5,6] and highlighted some of the difficulties in analyzing questions about local bulk physics in the AdS/CFT correspondence. Put briefly, these authors argued that the CFT does not contain operators with the right properties to play the role of local field operators behind the black hole horizon. Their arguments were phrased in terms of various paradoxes, and they interpreted these apparent contradictions to mean that generic high energy states in the CFT do not have a smooth interior; and even if they do, the CFT cannot describe it meaningfully.

If correct, this conclusion would be a striking violation of effective field theory. A semiclassical analysis performed by quantizing fluctuations about the classical black hole solution would suggest that for a large black hole, quantum effects detectable within effective field theory are confined to the neighborhood of the singularity. However, the papers above suggest that the range of quantum effects, visible to a low energy observer, may spread out all the way to the horizon.

In previous work [7–10], we analyzed these arguments in detail. We found that they made two tacit assumptions. The first, which was important for the strong subadditivity paradox of Mathur [1] and the first paper of Almheiri *et al.* [2], was that locality holds exactly in quantum gravity. We showed how a precise version of black hole complementarity, where the commutator of operators outside and inside the black hole vanishes within low point correlators but is not exactly zero as an operator, allow one to resolve this paradox. We review this resolution briefly at the end of Sec. VII D below.

We emphasize that this resolution is consistent with the belief that locality is not absolute in theories of quantum gravity; so a nonvanishing commutator between operators outside and inside is not surprising by itself. What we found, however, was that it was possible to construct interior operators so that this nonvanishing commutator only shows up in very delicate observations involving an extremely large number of quanta. The reader may wish to look at Sec. VII D and then at [9] for further discussion of these nonlocal effects.

Our focus in this paper is on a second aspect of the information paradox that was emphasized in [3]. Here, Almheiri *et al.* argued that even large black holes in anti–de Sitter (AdS) should contain firewalls. To make this argument they had to make a second tacit assumption, which was that local bulk observables like the metric are represented by fixed linear operators in the CFT. More precisely,

*kyriakos.papadodimas@cern.ch
†suvrat@icts.res.in

this is the idea that even in two different states one may use the same CFT operator to represent the metric at a "given point."

By identifying and discarding this assumption in [8,9], we were able to resolve all the paradoxes alluded to above. Furthermore, we were able to explicitly identify CFT observables that were dual to local correlation functions in the black hole interior. This construction allowed us to probe the geometry of the horizon and show that the horizon was smooth—as predicted by effective field theory, and in contradiction with the firewall and fuzzball proposals.

The operators in our construction are state dependent. This means that they act correctly about a given state, and in excitations produced on that state by performing low energy experiments. If one moves far in the Hilbert space—even just by changing the microscopic and not the macroscopic degrees of freedom—then one has to use a different operator to represent the "same" local degrees of freedom.

Our resolution to the firewall paradox has encountered two kinds of objections. A technical point is that our construction relies on a notion of equilibrium. It was first noticed by van Raamsdonk [11] that our equilibrium conditions were necessary but not sufficient; Harlow [12] later elaborated on this point. This leads to a potential "ambiguity" in our construction where, at times, we cannot definitively identify the right operators in the black hole interior.

The second is more fundamental. Is it acceptable at all, within quantum mechanics, to use state-dependent bulk to boundary maps so that the metric at a given point in space may be represented by different operators in different microstates and backgrounds? It has been argued [3,4,12] that state-dependence is inconsistent with linearity in quantum mechanics. Is this correct, and in particular, is it possible for any observer (bulk or boundary) to detect measurable violations of linearity?

This is the context for our paper. In this work we make the following advances.

(1) In Sec. V, we revisit and sharpen the arguments of Almheiri *et al.* [3]. We believe that this strongly suggests that there is no alternative to firewalls except for a state-dependent construction of the black hole interior. In fact, we show in Sec. VI that the paradoxes of [3] also arise for the eternal black hole. We show that it is necessary to use state-dependent operators, which we construct explicitly, to rule out a scenario where even the eternal black hole does not have a smooth interior.

(2) In Sec. VIII, we resolve a large class of ambiguities in our construction by refining our notion of an equilibrium state, including all of those pointed out by van Raamsdonk [11]. We point out difficulties with Harlow's analysis [12] that attempted to accentuate these ambiguities.

(3) We show how our analysis extends naturally to superpositions of states in Sec. VII. We reiterate and expand on the point, already made in [8,9] that the infalling observer does not observe any violations of quantum mechanics or the "Born rule."

(4) In Sec. IX, we show how our construction extends naturally to entangled systems. This leads to a new and interesting outcome: a precise version of the ER = EPR conjecture [13] emerges automatically from our analysis. In particular our construction shows—without any additional assumptions—why one should expect a geometric wormhole in the thermofield double state, and a somewhat "elongated" wormhole in states with less entanglement. Our analysis also shows why there is no geometric wormhole in a generic entangled state of two CFTs, or when the CFT is entangled with a system of a few qubits.[1]

We also initiate an investigation into the semiclassical origins of state-dependence in Appendix A. We show that local observables like the metric are well-defined classical functions on the phase space of canonical gravity. Ordinarily such functions would lift to state-independent operators in the quantum theory. However, our analysis of state-dependence in the eternal black hole suggests an interesting obstacle to this map: the inner product between states in the CFT representing different geometries does not die off as fast as a naive analysis of coherent states in canonical gravity would suggest. Instead it saturates at a nonperturbatively small but finite value. We present some evidence that it is this overcompleteness that prevents the existence of state-independent operators behind the horizon.[2]

Apart from the new results mentioned above, we also present some material that we hope will help to clarify some conceptual issues and be of pedagogical utility. For example, in Sec. III we present a discussion of relational observables in AdS quantum gravity. This concept is important throughout this paper to understand the geometric properties of operators behind the horizon, but we believe that it may be of broader significance. This idea has often been used in discussions of the subject (and was first described to us by Donald Marolf) but we attempt to present a pedagogical and precise definition here.

We also present a derivation of the properties of operators behind the horizon from a pedagogically new perspective in Sec. IV. We consider the two-point function of a massless scalar field propagating in the geometry. By using the properties of this two-point function, when the

[1]We limit our assertions to wormholes that can be probed geometrically using effective field theory. Therefore we do not have any comment on the strong form of the ER = EPR conjecture, which posits that any entanglement should be accompanied by a wormhole.

[2]A similar idea was suggested earlier by Motl [14].

two points are almost null to each other, we are able to derive the correct formula for the entanglement of modes behind and in front of the horizon. One concern about our previous analysis [7] was that even though a black hole in a single CFT does not have a second asymptotic region, we had to appeal to the analogy with the thermofield double, to derive the properties of our operators behind the horizon. We now perform this derivation from a purely local calculation.

We believe that the results of this paper present compelling evidence in favor of the claim that there are no firewalls in generic states, and also that the map between bulk and boundary operators is state dependent behind the horizon.

The recent literature on the information paradox is extensive [15]. In particular, Erik Verlinde and Herman Verlinde also reached the conclusion that state-dependence is required to construct the black hole interior from a different perspective [16,17]. We direct the reader to [18] for a discussion of the relation between our approach and theirs. The effects of the backreaction of Hawking radiation were discussed in [19], and Nomura *et al.* also presented another perspective in [20]. For a precursor of the firewall paradox, see [21] and for approaches using complexity see [22].

II. SUMMARY

In this section, we briefly summarize the contents of various sections and suggest different paths that could be taken through the paper.

Reconstructing the bulk and state-dependence: Section III is partly devoted to clarifying some conceptual issues related to bulk to boundary maps. We quickly review what it means for such a map to be state independent or state dependent. We also point out that all existing methods of extracting bulk physics from the boundary, as currently formulated, are state dependent. Experts in the subject may wish to look only at Sec. III A 1 where we define the relational observables that we use in the rest of the paper and at Sec. III B 1 where we describe the state-dependence of prescriptions to relate geometric quantities to entanglement.

Need for operators behind the horizon: Section IV is largely devoted to a detailed derivation of the fact that we require new modes that can play the role of "right moving" excitations behind the horizon to describe the interior of a black hole. We derive the two-point function of these modes with modes outside the horizon from a local calculation, thereby removing the need to make an analogy to the thermofield double state and also sidestepping the trans-Planckian issues in Hawking's original computation. In this section, we also review the standard construction of local operators outside the horizon. Experts may be interested in Sec. IV B 2 where we describe a state-independent construction of local operators outside the horizon in the minisuperspace approximation.

Either state-dependence or firewalls: The objective of Sec. V is to try and show that we must accept one of two possibilities: either the black hole interior is mapped to the CFT by a state-dependent map, or generic microstates have firewalls. Our arguments here are extensions and refinements of the arguments presented in [3,4]. In particular, we strengthen the argument of [4] by bounding potential errors in that calculation. We also rephrase the "counting argument" of [3] entirely within the context of two-point correlation functions to remove potential loopholes. This section can be skipped, at a first reading, by a reader who already accepts the validity of the arguments of [3,4].

State-dependence for the eternal black hole: In Sec. VI we show that these versions of the information paradox also appear in the eternal black hole. Therefore it is inconsistent to adopt the position that the eternal black hole in AdS has a smooth interior whereas the large single-sided black hole does not. We would urge the reader to consult [23]—where a concise version of these arguments has already appeared—in conjunction with this section, which contains some additional details. Since there is substantial evidence that the interior of the eternal black hole is smooth, this provides strong support for state-dependence behind the black hole horizon.

Definition of mirror operators; consistency with superposition principle: In Sec. VII, we review the state-dependent construction of the black hole interior that was first presented in [8,9]. Experts may be interested in Sec. VII E where we check the linearity of this map for superpositions of a small number of states. In Sec. VII F we construct the interior of the eternal black hole. This construction is of interest since it provides some insight into state-dependence as arising from the "fat tail" of the inner product between different microstates of a black hole.

Detecting unitaries behind the horizon: In Sec. VIII, we show how to remove some of the ambiguities in our definition of equilibrium. This section will be of interest to experts. We point out that by using the CFT Hamiltonian, we can detect excitations behind the horizon in states that we might otherwise have classified as being in equilibrium. We also point out, in some detail, that the effort made in [12] to sharpen this ambiguity by considering a new class of excitations is based on an erroneous analysis of local operators in the eternal black hole. While, for this reason, the analysis of [12] does not have direct physical significance, it does point to an interesting new class of excited states that we discuss in some detail.

Entangled systems and relation to ER = EPR: In Sec. IX, we extend our construction to account for cases where the CFT is entangled with another system. The equations that describe modes in the interior do not change at all. The only new element that we need to introduce is that the "little Hilbert space" of excitations about a base state may get enlarged since we can also act with operators in the other system. Surprisingly we show that a precise version of the ER = EPR conjecture emerges automatically from our analysis. We are able to show that when two

entangled CFTs are in the thermofield state the modes observed by the right-infalling observer inside the black hole are the same as those observed by the left observer outside. However, when the CFTs are entangled in a generic manner this is no longer true.

We also consider cases where the CFT is entangled with a small system—say a collection of qubits. Our analysis of this setup, together with our verification of linearity in Sec. VII establishes that the infalling observer cannot detect any departures from ordinary linear quantum mechanics.

III. GENERALITIES: STATE-DEPENDENT VS STATE-INDEPENDENT OPERATORS

Since this paper focuses on state-dependent bulk-boundary maps, it is useful to first clarify the meaning of state-dependence and, conversely, what we would require of a putative state-independent operator. Since this issue has been the cause of significant confusion—some of which has arisen because of the use of imprecise terminology—we have tried to make this section as precise and detailed as possible.

A brief summary of this section is as follows. We define state-dependence. We point out that state-dependent bulk-boundary maps are already common in the AdS/CFT literature. Finally we explain the origin of the naive expectation that the bulk and boundary are related in a state-independent manner, and also indicate why this intuition fails.

Apart from the pedagogical definitions, we also pay some attention to the techniques of extracting bulk physics using entanglement entropy. These are all state dependent since entanglement entropy does not correspond to a linear operator on the boundary. This includes, for example, the well-known Ryu-Takayanagi (RT) relation [24] between the entanglement entropy of a region on the boundary and the corresponding area of an extremal surface in the bulk. As we emphasize repeatedly in this paper, as a result of very robust statistical properties of the Hilbert space of the CFT at large $\mathcal{N}$,[3] it is perfectly natural for such a state-dependent formula to emerge within effective field theory, and its use does not lead to any violation of quantum mechanics.

While the use of state-dependent operators may be common in AdS/CFT, from a broader viewpoint it is true that this is a rather special situation in physics. So it would be incorrect to go to the other extreme and dismiss state-dependence as mundane or unremarkable.

In this section, we point out that based on intuition from canonical gravity, one may have naively expected that there is some overarching linear operator in the CFT that includes, in various limits, all these state-dependent prescriptions. If one were to obtain gravity through phase space quantization, then one may naively expect that many reasonable functions on the phase space of gravity—such as the metric at a point—would lift to operators. We show why this naive intuition runs into difficulty in the context of AdS/CFT. We complete this analysis in greater detail in Appendix A. The semiclassical origins of state-dependence that we outline in this section and in the appendix are, we believe, an important and interesting subject of study.

In this section and later in the paper we often speak of CFT operators that also have a dual geometric interpretation. To avoid confusion, we adopt the following notational convention.

Notation: A CFT operator is denoted with a bold symbol; for example an operator in the CFT corresponding to the bulk metric would be denoted by $\boldsymbol{g}_{\mu\nu}$, as opposed to the value of the semiclassical metric for a geometry $g_{\mu\nu}$, which is written in ordinary font.

A. State-independent operators

We consider an AdS/CFT duality, where we expect a number of "effective fields" to propagate in the bulk. One of these is the metric $\boldsymbol{g}_{\mu\nu}$ but in general there are other fields, which can include scalars but also fields of higher spin. We collectively denote these fields by $\boldsymbol{\phi}$. We then have the following definition.

Definition of a state-independent bulk-boundary map: We say that there is a state-independent map between the bulk and the boundary if there exist CFT operators $\boldsymbol{g}_{\mu\nu}(\vec{x})$ and $\boldsymbol{\phi}(\vec{x})$ parametrized by $d+1$ real numbers, which we denote by $\vec{x}$, so that in *all CFT* states that are expected to be dual to a semiclassical geometry, which we denote by $|\Psi\rangle$, the CFT correlators involving both the metric and other light fields,

$$C(\vec{x}_1,\ldots\vec{x}_{m+p}) = \langle\Psi|\boldsymbol{g}_{\mu_1\nu_1}(\vec{x}_1)\ldots\boldsymbol{g}_{\mu_m\nu_m}(\vec{x}_m)\boldsymbol{\phi}(\vec{x}_{m+1})\ldots\boldsymbol{\phi}(\vec{x}_{m+p})|\Psi\rangle, \tag{3.1}$$

have the right properties to be interpreted as "effective field theory correlators."

This definition has many parts that we unpack below, where we explain what it means for a state to be dual to a semiclassical geometry, and what one expects from effective field theory.

An immediate issue—but one that does not have significant physical ramifications—is that the bulk theory has diffeomorphism invariance. The $d+1$ real numbers above play the role of coordinates in the bulk. Given any valid diffeomorphism, $\vec{x}\to\xi(\vec{x})$, the *distinct* CFT operators $\boldsymbol{\phi}(\xi^{-1}(\vec{x}))$ give an equally valid bulk to boundary map. So we must always discuss equivalence classes of bulk-boundary maps. Maps that are related by diffeomorphisms

[3]In this paper we adopt notation that is consistent with [8,9]. So $\mathcal{N}$ is proportional to the central charge of the CFT. In the commonly considered case of the maximally supersymmetric SU(N) Yang-Mills theory, we would have $\mathcal{N}\propto N^2$.

belong to the same equivalence class. Later in this section, we also describe various physical choices of gauge that help to remove this redundancy, and pick a preferred element of the equivalence class. We now turn to other aspects of the definition above.

Semiclassical states: We now explain what we mean by semiclassical states in the definition above. In the AdS/CFT duality, we often identify certain states with dual bulk geometries. These maps have been developed as a result of various calculations. Schematically, we may represent this process of identifying a metric dual to a state by

$$|\Psi_g\rangle \leftrightarrow g_{\mu\nu}(\vec{x}). \tag{3.2}$$

Two examples may help in elucidating this concept. Consider the vacuum of the CFT, $|0\rangle$. In this case, the expectation is that

$$|0\rangle \leftrightarrow g^{\rm ads}_{\mu\nu},$$

where the metric on the right-hand side is the metric of *empty global AdS*.

In this paper, we are particularly interested in a second example of such maps: a generic state at high energies in the CFT is believed to be dual to a large black hole in the bulk.

Consider a set of energy eigenstates centered around a high energy $E_0 \gg \mathcal{N}$, and with a width $\Delta \ll \mathcal{N}$. The set of all energy eigenstates in this range is called

$$\mathcal{R}_{E_0} \equiv \{|E_i\rangle : E_0 - \Delta \le E_i \le E_0 + \Delta\}.$$

We denote the dimension of this space by $\mathcal{D}_{E_0}$. By taking all linear combinations of these states, we get a subspace of the Hilbert space of the CFT,

$$|\Psi\rangle = \sum \alpha_i |E_i\rangle, \qquad |E_i\rangle \in \mathcal{R}_{E0}. \tag{3.3}$$

We assume above (and whenever we use α_i to take superpositions of states) that they are chosen so that the state is correctly normalized. We can place an additional restriction on $|\Psi\rangle$ above that it has vanishing $SO(d)$ and R-charges.

Next, we consider the set of unitary matrices that acts entirely within this subspace. This is a very large unitary group $U(\mathcal{D}_{E_0})$. For $\Delta = \mathrm{O}(1)$, we expect that $\mathcal{D}_{E_0} = \mathrm{O}(e^{\mathcal{N}})$. The Haar measure on this unitary group now defines a measure for the coefficients α_i in (3.3), and we can pick a "typical" state in the microcanonical ensemble by using this measure. Then the expectation is that almost all states chosen in this manner, except for an exponentially small fraction of states, correspond to a dual Schwarzschild black hole geometry in the bulk:

$$|\Psi\rangle \leftrightarrow g^{\rm bh}_{\mu\nu}.$$

We can get other kinds of black holes by varying the other charges. This is the central class of "semiclassical states" that we are interested in this paper.

The example above also points to an additional important fact, which the reader should keep in mind. While we write $|\Psi_g\rangle$ to prevent the notation from becoming unwieldy the state dual to a geometry is far from unique. There are several microstates that represent the same geometry.

Two additional classes of states are of some interest to us, and are entirely derivative from the class above.

(1) Superpositions of semiclassical states

First, given states corresponding to different metrics $|\Psi_{g_1}\rangle \leftrightarrow g_{1,\mu\nu}, \ldots, |\Psi_{g_m}\rangle \leftrightarrow g_{m,\mu\nu}$ we may consider a superposition of such states,

$$|\Psi_s\rangle \equiv \left(\sum_{i=1}^{m} \alpha_i |\Psi_{g_i}\rangle\right). \tag{3.4}$$

If the geometries above are reasonably distinct, then the states are almost orthogonal. This is also the case if we pick two generic microstates corresponding to the *same geometry*. As we see below we expect that

$$\begin{aligned} &\langle \Psi_{g_1} | \Psi_{g_2} \rangle = \mathrm{O}(e^{-\mathcal{N}}), \\ &\langle \Psi_{g_1} | \boldsymbol{g}_{\mu_1\nu_1}(\vec{x}_1) \ldots \boldsymbol{g}_{\mu_m\nu_m}(\vec{x}_m) \boldsymbol{\phi}(\vec{x}_{m+1}) \ldots \boldsymbol{\phi}(\vec{x}_{m+p}) | \Psi_{g_2} \rangle \\ &\quad = \mathrm{O}(e^{-\mathcal{N}}), \end{aligned} \tag{3.5}$$

both for states corresponding to distinct geometries, and for generic microstates corresponding to the same geometry. Therefore, we require $\sum_i |\alpha_i|^2 = 1 + \mathrm{O}(e^{-\mathcal{N}})$ in this situation. The important point is as follows. The smallness of the off-diagonal matrix elements above implies that a quantum superposition of a small number of geometries, or a small number of microstates corresponding to the same geometry, corresponds in effect to a classical probability distribution over these states. On the other hand, it is clear that if we take $m = \mathrm{O}(e^{\mathcal{N}})$ in the superposition above, then this intuition breaks down, and the cross terms become important.

(2) Excitations of semiclassical states

Furthermore, given a state $|\Psi_g\rangle$, which we have identified with a metric $g_{\mu\nu}$, one can consider "excitations" of this state. For example, one may "act" on this state using some of the operators corresponding to the metric or other light fields. These new states correspond to excitations of the original state,

$$\begin{aligned} |\Psi^{\rm ex}_g\rangle = \boldsymbol{g}_{\mu_1\nu_1}(\vec{x}_1) \ldots \boldsymbol{g}_{\mu_m\nu_m}(\vec{x}_m) \ldots \boldsymbol{\phi}(\vec{x}_{m+1}) \ldots \\ \boldsymbol{\phi}(\vec{x}_{m+n}) |\Psi_g\rangle. \end{aligned} \tag{3.6}$$

In the large $\mathcal{N}$ limit, after subtracting off the contribution of the background metric, this state

should be interpreted as an excitation with $n+m \ll \mathcal{N}$ quanta on a background with metric g. Although these excited states occupy a very small fraction of the volume of the Hilbert space at any energy, they are important because there are several interesting physical questions about the response of equilibrium states to excitations.

Coherent states vs metric eigenstates: Although we have taken a CFT perspective on the states above, in principle we could also have viewed these states as solutions of the Wheeler de Witt equation that live in a Hilbert space obtained by quantizing gravity and the other light fields. From this perspective we should emphasize, to avoid any confusion, that the semiclassical states $|\Psi_g\rangle$ that we refer to here are "coherent states," which correspond to an entire semiclassical spacetime; these states are *distinct* from metric eigenstates that are sometimes considered in conventional analyses of canonical gravity.[4]

Let us make this more precise. We start by performing a $d+1$ split of the geometry

$$ds^2 = -N^2dt^2 + \gamma_{ij}(dx^i + N^i dt)(dx^j + N^j dt),$$

and promote the d-metric γ_{ij} to an operator. The canonically conjugate momentum is

$$\pi^{ij} = -\gamma^{\frac{1}{2}}(K^{ij} - \gamma^{ij}K),$$

where K^{ij} is the extrinsic curvature [25]. [See (A6) for an explicit expression.] Given a CFT operator $\boldsymbol{g}_{\mu\nu}$ we can therefore define two related CFT operators $\boldsymbol{\gamma}_{ij}$ and $\boldsymbol{\pi}^{ij}$. Now the key point is that the semiclassical/coherent states that we are discussing satisfy

$$\begin{aligned}
&\langle\Psi_g|\boldsymbol{\gamma}_{i_1j_1}(\vec{x}_1)\boldsymbol{\gamma}_{i_2j_2}(\vec{x}_2)|\Psi_g\rangle \\
&\quad - \langle\Psi_g|\boldsymbol{\gamma}_{i_1j_1}(\vec{x}_1)|\Psi_g\rangle\langle\Psi_g|\boldsymbol{\gamma}_{i_2j_2}(\vec{x}_2)|\Psi_g\rangle = \mathrm{O}\left(\frac{1}{\mathcal{N}}\right), \\
&\langle\Psi_g|\boldsymbol{\pi}^{i_1j_1}(\vec{x}_1)\boldsymbol{\pi}^{i_2j_2}(\vec{x}_2)|\Psi_g\rangle \\
&\quad - \langle\Psi_g|\boldsymbol{\pi}^{i_1j_1}(\vec{x}_1)|\Psi_g\rangle\langle\Psi_g|\boldsymbol{\pi}^{i_2j_2}(\vec{x}_2)|\Psi_g\rangle = \mathrm{O}\left(\frac{1}{\mathcal{N}}\right).
\end{aligned} \tag{3.7}$$

We can specify the $\mathrm{O}(\frac{1}{\mathcal{N}})$ terms precisely, as we do in the next section. But for now we emphasize that these states have a small but finite uncertainty for both the three-metric and its canonically conjugate variable. Therefore they are distinct from metric eigenstates which would have satisfied

$$\boldsymbol{\gamma}_{ij}(\vec{x})|\gamma\rangle = \gamma_{ij}(\vec{x})|\gamma\rangle, \qquad \text{metric eigenstate.}$$

Such metric eigenstates would, on the other hand, have a large variance for $\boldsymbol{\pi}^{ij}$.

It is these coherent states that have a natural semiclassical interpretation. Metric eigenstates, on the other hand, have maximum uncertainty in the value of $\boldsymbol{\pi}^{ij}$ and therefore, under time evolution, they quickly disperse into a superposition of several different eigenstates.

Expectations from effective field theory: We now turn to the other term used in the definition above: the expectations from effective field theory for correlators of these operators.

Let us assume that we are given a state $|\Psi_g\rangle$ which is believed to be dual to a geometry by the relation (3.2). Then, the most basic expectation from a putative CFT operator that could yield the metric in the bulk is that

$$\langle\Psi_g|\boldsymbol{g}_{\mu\nu}(\vec{x})|\Psi_g\rangle = g_{\mu\nu}(\vec{x}). \tag{3.8}$$

Further, we demand that the n-point correlators of these operators have the property that

$$\begin{aligned}
&\langle\Psi_g|\boldsymbol{g}_{\mu_1\nu_1}(\vec{x}_1)\ldots\boldsymbol{g}_{\mu_n\nu_n}(\vec{x}_n)|\Psi_g\rangle \\
&\quad = g_{\mu_1\nu_1}(\vec{x}_1)g_{\mu_2\nu_2}(\vec{x}_2)\ldots g_{\mu_n\nu_n}(\vec{x}_n) \\
&\qquad + G_{\mu_1\nu_1\mu_2\nu_2}(\vec{x}_1,\vec{x}_2)g_{\mu_3\nu_3}(\vec{x}_3)\ldots g_{\mu_n\nu_n}(\vec{x}_n) + \text{perm} \\
&\qquad + G_{\mu_1\nu_1\mu_2\nu_2\mu_3\nu_3}(\vec{x}_1,\vec{x}_2,\vec{x}_3)g_{\mu_4\nu_4}(\vec{x}_4)\ldots g_{\mu_n\nu_n}(\vec{x}_n) + \text{perm} \\
&\qquad + \cdots.
\end{aligned} \tag{3.9}$$

where $G_{\mu_1\nu_1\ldots\mu_j\nu_j}(\vec{x}_1,\ldots\vec{x}_j)$ are the *connected j-point correlators* as calculated by perturbatively quantizing metric fluctuations on the background of the metric $g_{\mu\nu}$ and … are the higher point functions which we have not shown explicitly. Note that this also fixes the $\frac{1}{\mathcal{N}}$ corrections that appeared in (3.7), because the connected correlators are subleading in $\frac{1}{\mathcal{N}}$.

Similarly, we declare that other bulk excitations are realized by state-independent operators, if there exist operators $\boldsymbol{\phi}(\vec{x})$ in the CFT, with the property that n-point correlators of these operators have an expansion

$$\begin{aligned}
&\langle\Psi_g|\boldsymbol{\phi}(\vec{x}_1)\boldsymbol{\phi}(\vec{x}_2)\ldots\boldsymbol{\phi}(\vec{x}_n)|\Psi_g\rangle \\
&\quad = G(\vec{x}_1,\vec{x}_2)G(\vec{x}_3,\vec{x}_4)\ldots G(\vec{x}_{n-1},\vec{x}_n) + \text{perm} \\
&\qquad + G(\vec{x}_1,\vec{x}_2,\vec{x}_3)G(\vec{x}_4,\vec{x}_5,\vec{x}_6)G(\vec{x}_7,\vec{x}_8)\ldots G(\vec{x}_{n-1},\vec{x}_n) \\
&\qquad + \text{perm} + \cdots,
\end{aligned} \tag{3.10}$$

where the functions G are the perturbative j-point connected correlation functions as obtained by quantizing the field ϕ about the metric g.

[4]Strictly speaking, if we think of the degrees of freedom in gravity as being obtained from tracing out stringy and other heavy degrees of freedom, then we would expect a generic CFT state to correspond to a density matrix for the gravitational degrees of freedom, and not a pure state at all. However, because off-diagonal matrix elements of light operators between different coherent states are very small, a sum of coherent states effectively behaves like a classical superposition. Therefore we can neglect this complication here. Indeed, it is because of this fact that canonical gravity—where the entanglement with these heavier degrees of freedom is ignored even in excited background geometries like the black hole—makes sense at all.

In this expansion, we emphasize that we are not interested in gravitational loop corrections at the moment, but would be satisfied if the n-point correlators of the CFT operators have an expansion that agrees with that obtained from perturbative quantum field theory carried out at tree level. This tree-level contribution is already enough to fix the leading $\frac{1}{\mathcal{N}}$ terms. It is also important to note that even the two-point function already knows about the background metric. This is simply because the graviton and matter propagators depend on the metric background. Therefore, in a sense, in the expansions (3.9)–(3.10) we have already resummed the $\frac{1}{\mathcal{N}}$ series. It is in this resummed series that we are only interested in tree-level correlators.

Second, let us make a comment about superpositions of distinct geometries as in (3.4). Then we expect that

$$\langle \Psi_s | \boldsymbol{g}_{\mu\nu}(\vec{x}) | \Psi_s \rangle = \sum_{i=1}^{m} |\alpha_i|^2 \langle \Psi_{g_i} | g_{\mu\nu}(\vec{x}) | \Psi_{g_i} \rangle + \mathrm{O}(e^{-\mathcal{N}}).$$

A similar relation holds for n-point correlators, provided that $n \ll \mathcal{N}$. This is the statement that cross terms between macroscopically distinct geometries are very small. So, a superposition of the form above essentially behaves like a classical mixture for our purposes.

This is an important point since there is no canonical way to speak of the "same point" in different macroscopic geometries. Stated precisely, this is the statement that quantum field theory in curved spacetime does not lead to any prediction for cross-correlators

$$\langle \Psi_g | \boldsymbol{g}_{\mu\nu}(\vec{x}_1) \boldsymbol{g}_{\mu\nu}(\vec{x}_2) | \Psi_{g'} \rangle,$$

where $g_{\mu\nu}(\vec{x})$ and $g'_{\mu\nu}(\vec{x})$ are metrics corresponding to macroscopically different geometries.[5] However, (3.5) tells us that we *never* need to consider such cross terms in correlators of the metric, which are exponentially suppressed and do not have any semiclassical interpretation.

Finally, let us point out that if we declare that we do have a construction of state-independent local operators, then we should take it seriously. Therefore, if we find a state $|\Psi\rangle$, in which n-point correlators of the operator $\boldsymbol{\phi}(\vec{x})$ cannot be reorganized as perturbative correlators about any metric, then we must declare that the state $|\Psi\rangle$ does not correspond to a semiclassical geometry.

Gauge invariance and coordinates: We now turn to the last remaining point in our definition of state-independent operators. The $d+1$ real parameters parametrize CFT operators and are to be interpreted as coordinates in AdS. This is a tractable issue but two points sometimes lead to confusion: the fact that the metric and other local observables are not gauge invariant, and the fact that we are using a uniform coordinate system to represent all metrics. Both of these issues can be resolved simultaneously by an appropriate gauge fixing, as we now describe.

First, as we have already noted, given a family of CFT operators labeled by coordinates $\vec{x}$, so that the family of operators satisfies (3.8)–(3.9) we can clearly simply consider another family of CFT operators, which is related to the previous one by diffeomorphisms.

$$\bar{\boldsymbol{g}}^{\mu\nu}(\vec{x}) = \frac{\partial \xi^\mu}{\partial x^\rho} \frac{\partial \xi^\nu}{\partial x^\sigma} \boldsymbol{g}^{\rho\sigma}(\vec{\xi}^{-1}(\vec{x})),$$
$$\bar{\boldsymbol{\phi}}(\vec{x}) = \boldsymbol{\phi}(\xi^{-1}(\vec{x})). \tag{3.11}$$

The operators on the left-hand side of (3.11) are distinct CFT operators, but they obviously encode the same bulk physics. We can choose to simply live with this lack of uniqueness, while keeping in mind that to extract any physics from the operator (3.8) we need to form gauge-invariant quantities. But from a physical point of view, it is more convenient to pick a gauge so that the CFT operators that we are discussing become unambiguous.

A related problem has to do with the "range" of the real numbers in $\vec{x}$. Usually, we tailor the coordinate system to the metric. So it is often the case that the AdS Schwarzschild metric and the empty AdS metric are written in terms of coordinates that have different ranges.

In addressing these two issues, it is useful to recognize that they also arise in numerical general relativity. There we are given a grid of points, drawn from $R^{d,1}$, with a fixed range and we place different metrics on this grid so that the resultant spacetime describes an entire range of physics, from empty AdS to black holes.

To make this more precise, note that the empty AdS metric is given by

$$ds^2_{\mathrm{ads}} = -(1+r^2)dt^2 + \frac{dr^2}{1+r^2} + r^2 d\Omega^2_{d-1}.$$

By a coordinate transformation, $r = \frac{\rho}{1-\rho}$, we can bring the boundary to a finite coordinate distance

$$ds^2_{\mathrm{ads}} = \frac{1}{(1-\rho)^2}\left(-\hat{f}(\rho)dt^2 + \frac{1}{\hat{f}(\rho)}d\rho^2 + \rho^2 d\Omega^2_{d-1}\right), \tag{3.12}$$

with $\hat{f}(\rho) = (1-\rho)^2 + \rho^2$. The boundary is at $\rho = 1$, and manifold in (3.12) is $[0,1) \times R \times S^{d-1}$. In this paper we are only interested in different metrics placed on this manifold that asymptotically tend to the metric in (3.12), although they may differ in the bulk. Even if black holes are present, we simply consider nice slices that are parametrized by the coordinates $[0,1) \times S^{d-1}$, as shown in Fig. 1, and then consider their evolution in time for a finite range of time. Note that by this finite-time restriction, we also avoid questions of "topology changes."

[5]For the case where these metrics are so close that one can be considered to be a coherent excitation of gravitons on the other, we refer the reader to Appendix A.

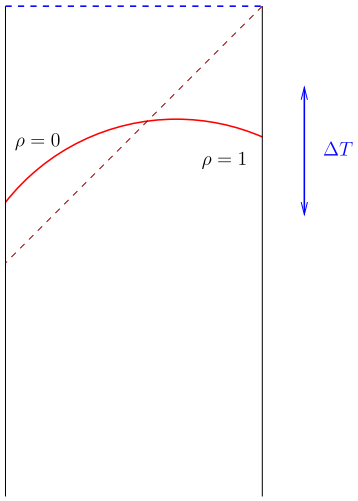


FIG. 1. Even in the presence of a black hole, nice slices can be parametrized by coordinates on $[0,1) \times S^{d-1}$. We examine physics for a finite interval ΔT so that the future singularity is irrelevant.

Having chosen a uniform coordinate system to describe the metrics that we are interested in, we can further choose a gauge to unambiguously specify the CFT operators we are interested in. A convenient choice of gauge is given by the "generalized harmonic gauges." In these gauges, we set

$$\Box \vec{x}^{\mu} = H^{\mu}(\vec{x}). \tag{3.13}$$

A choice of the "source functions" $H^{\mu}(\vec{x})$ gives a choice of gauge.

Note that once (3.13) is imposed as an additional *operator equation* that must be satisfied by the CFT operators that appear in (3.8)–(3.9), then this removes the redundancy (3.11) in the identification of these operators in the CFT. So, if such operators exist then (3.13) picks out a specific family of them.

For the specific case of AdS, an appropriate choice of source functions is discussed in detail in [26]. These details are not important here. The point that we can take away from the numerical analysis of [26] is that it is possible to describe a very broad range of metrics in AdS, including empty AdS and excited black holes that are dual to fluid dynamical situations on the boundary with a uniform choice of coordinate system and gauge.

1. Relational observables

There is another class of coordinate systems, which is particularly convenient in AdS. This is the class of coordinate systems that is defined relationally with respect to the boundary. Here, we assume that we are already given the metric in some coordinate system, such as the ones above. We then describe a coordinate transformation to a more *convenient* relational coordinate system.

Intuitively, we consider an experiment where an observer jumps from the boundary, with no initial velocity along the S^{d-1}, falls for a given amount of proper time, and then makes a measurement. In fact this notion is a little hard to make concrete in this form because if we drop the observer from a point that is infinitesimally close to the boundary, he very rapidly approaches the speed of light. This problem cannot be solved by using an affine parametrization of null geodesics either, since any affine parameter that is finite in the bulk goes to infinity as we reach the boundary.

So, it is convenient to use the following slightly more complicated construction. We start from a given point on the boundary, which we label by (t_1, Ω_1). We know that the metric is of the asymptotically AdS form given by (3.12). We now consider a null geodesic, parametrized by ordinary asymptotic AdS time, that extends into the bulk, with no velocity along the S^{d-1}. More precisely, let us consider a null geodesic trajectory given by

$$\begin{aligned} &\vec{x}_1(t) \equiv (t, \rho_1(t), \quad \Omega_1(t)), \quad \rho_1(t_1) = 1, \\ &\Omega_1(t_1) = \Omega_1, \quad \dot{\Omega}_1(t_1) = 0, \quad \dot{\rho}(t_1) = -1, \end{aligned} \tag{3.14}$$

where by a slight abuse of notation we have used Ω_1 both for the solution to the geodesic equation, and for the initial value of the solution. Note that initial "velocity" in the radial direction is fixed since the geodesic is null and the sign indicates that the geodesic is ingoing and moves into the bulk as time advances. This geodesic reaches a finite coordinate distance in the bulk in finite time. Second, note that while we are starting with no angular momentum, intrinsic properties of the geometry may cause the geodesic to start moving on the sphere as well after it departs from the boundary.

We now consider a *second* null geodesic that intersects the boundary at a *later* point $(t_1 + \tau, \Omega_2)$ and also has $\dot{\Omega}_2 = 0$ at its final point. This is the geodesic trajectory

$$\begin{aligned} \vec{x}_2(t) &= (t, \Omega_2(t), \rho_2(t)), \\ \rho_2(t_1 + \tau) &= 1, \\ \Omega_2(t_1 + \tau) &= \Omega_2, \\ \dot{\Omega}_2(t_1 + \tau) &= 0 \\ \dot{\rho}_2(t_1 + \tau) &= 1, \end{aligned} \tag{3.15}$$

and the sign of the radial derivative indicates that the geodesic is outgoing at the time $t_1 + \tau$. Now given a particular value of $t_1, \Omega_1(t_1)$, we vary $\Omega_2(t_1 + \tau)$ so that the geodesics intersect. We expect that

$$\begin{aligned} &\exists \Omega_2 \quad \text{and} \quad \exists t_i, \\ &t_1 < t_i < t_1 + \tau \quad \text{such that } \rho_2(t_i) = \rho_1(t_i); \\ &\Omega_2(t_i) = \Omega_1(t_i). \end{aligned}$$

Intuitively, the existence of such a solution seems clear. For example, in the case where the geometry has no angular momentum at all, we can solve the equation above simply by setting $\Omega_2 = \Omega_1$. If we start deforming the geometry so that it is rotating, we should still be able to tune Ω_2 so that the two geodesics intersect. Even for other, more complicated geometries, we expect that the intersection point should be well defined at least as long as we are close enough to the boundary and we see below that this is all that we need.

We denote the point of intersection by

$$P_i(t_1, \Omega_1, \tau) \equiv (t_i, \Omega_1(t_i), \rho_1(t_i)). \tag{3.16}$$

This is a bulk point that is parametrized by the starting point of the first geodesic and the time difference to the ending point of the second geodesic.

Note that by means of such a process we cannot reach behind the black hole horizon. However, once we have a parametrization of points in the exterior, it is simple to extend them behind the horizon. We once again consider geodesics that start from a point $(t_1, \vec{\Omega}_1)$ on the boundary but this time we parametrize them using an affine parameter so that the geodesic satisfies the equation

$$\frac{d^2 x_1^\mu(\lambda)}{d\lambda^2} + \Gamma^\mu_{\nu\sigma} \frac{dx_1^\nu(\lambda)}{d\lambda} \frac{dx_1^\sigma(\lambda)}{d\lambda} = 0.$$

This is just a reparametrization of the geodesic in (3.14), and so we have denoted it with the same symbol $\vec{x}_1(\lambda)$.

The key point is that we can use our previous parametrization (3.16) to normalize the affine parameter. We set

$$\vec{x}_1(0) = P_i(t_1, \Omega_1, \tau_1), \qquad \vec{x}_1(1) = P_i(t_1, \Omega_1, \tau_2).$$

A choice of the intervals τ_1, τ_2 gives a specific normalization of the affine parameter. The reader can, for her convenience, think of any concrete value: say $\tau_1 = \ell_{\text{ads}}$, $\tau_2 = 2\ell_{\text{ads}}$.

Once this normalization is fixed we obtain the set of points

$$P_\lambda(t_1, \Omega_1, \lambda) = (t_1(\lambda), \Omega_1(\lambda), \rho_1(\lambda)). \tag{3.17}$$

The difference between (3.17) and (3.16) is that the points in (3.17) can also reach inside the horizon. The entire process above is summarized in Fig 2.

The advantage of this prescription is that, classically, measurements of a scalar field defined in such a relational manner are gauge invariant. We recall that when we define quantum gravity in anti–de Sitter space, we have to consider the set of all field configurations modulo trivial diffeomorphisms. The trivial diffeomorphisms are those that vanish at the boundary of anti–de Sitter space. Large gauge transformations—which leave the boundary in asymptotically AdS form, but yet move points on the boundary—correspond to symmetries in the boundary theory, and induce a change of the physical state.

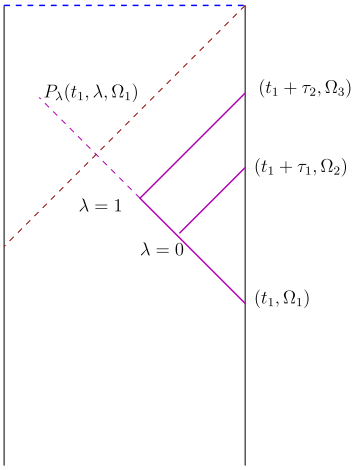


FIG. 2. The relational gauge fixing proceeds in two steps: first we use intersecting geodesics to parametrize points outside the horizon. Then we use this set of points to normalize the affine parameter and follow null geodesics into the horizon.

So, *gauge-invariant* observables are those that are invariant under trivial diffeomorphisms. In the relational observables described above, we start with a point on the boundary—which is left fixed because the diffeomorphism vanishes there—and then follow a gauge-invariant prescription to reach a point in the interior. Evidently, scalar fields evaluated at this point are themselves gauge invariant.

There is an important stronger statement that we can make. Consider a large diffeomorphism that induces a conformal transformation on the boundary $(t, \Omega) \to \mathcal{C}^{-1}(t, \Omega)$, where $\mathcal{C}$ denotes an element of the conformal group. Geometrically, under the diffeomorphism the geodesic trajectories in (3.14)–(3.15) get mapped to new geodesic trajectories. Therefore we expect that the relationally defined points in (3.17) will transform under the diffeomorphism as

$$P_\lambda(t, \Omega, \lambda) \to P_\lambda(\mathcal{C}^{-1}(t, \Omega), \lambda).$$

The important point is that this transformation of the relational points does *not* depend on the details of the diffeomorphism in the bulk, but merely on how it acts on the boundary.

Now consider a scalar field operator $\boldsymbol{\phi}(P_\lambda(t, \Omega), \lambda)$ with the bulk point defined as in (3.17). Corresponding to the conformal transformation $\mathcal{C}$, there is a unitary operator $U_\mathcal{C}$ on the boundary. Then, in order to be consistent with the geometric intuition, we expect that the CFT operator $\boldsymbol{\phi}$ will satisfy

$$U_\mathcal{C}^\dagger \boldsymbol{\phi}^i(P_\lambda(t, \Omega, \lambda)) U_\mathcal{C} = \boldsymbol{\phi}^i(P_\lambda(\mathcal{C}(t, \Omega), \lambda).$$

We use this relation several times to obtain the commutator of bulk operators with the Hamiltonian which arises

from the special case where $\mathcal{C}$ is just taken to be the time translation above. In Sec. VI, we apply this analysis in a more general setting where there are two boundaries.

The disadvantage of the relational prescription is that it is harder to make this precise at subleading order in $\frac{1}{\mathcal{N}}$. Clearly, the affine parameter along a geodesic from the boundary to another point may itself be expected to fluctuate at order $\frac{1}{\mathcal{N}}$. In this paper, these subtleties are not important.

B. The alternative: state-dependent bulk-boundary maps

An alternative to the state-independent possibility above is that geometric quantities like the metric do not arise by evaluating a Hermitian operator, but correspond to more general "measurables." More precisely, we would be led to state-dependence if there are no globally defined Hermitian operators $\boldsymbol{g}_{\mu\nu}(\vec{x})$ and $\boldsymbol{\phi}(\vec{x})$. Rather, about a given state $|\Psi_g\rangle$ we would have operators $\boldsymbol{g}_{\mu\nu}^{\{\Psi\}}(\vec{x})$ and $\boldsymbol{\phi}^{\{\Psi\}}(\vec{x})$ so that the correlators

$$C_\Psi(\vec{x}_1,\ldots\vec{x}_{m+p}) = \langle\Psi|\boldsymbol{g}_{\mu_1\nu_1}^{\{\Psi\}}(\vec{x}_1)\ldots\boldsymbol{g}_{\mu_m\nu_m}^{\{\Psi\}}(\vec{x}_m)\boldsymbol{\phi}^{\{\Psi\}}(\vec{x}_{m+1})\ldots\boldsymbol{\phi}^{\{\Psi\}}(\vec{x}_{m+p})|\Psi\rangle \tag{3.18}$$

reproduce the predictions of effective field theory that we outlined above. This definition is identical to the definition (3.1) in terms of the semiclassical states $|\Psi\rangle$ that appear here and the expectations we have for the values of the correlators. The difference is in the nature of the operators $\boldsymbol{g}_{\mu\nu}^{\{\psi\}}$ which now depend on the state.

One possible way to think about (3.18) is that the geometry emerges as a "function of correlation functions"[6] and not by measuring linear operators. However, we have some additional structure in (3.18). Since the bulk observer must see *quantum effective field theory*, it must be the case that to an excellent approximation the operators $\boldsymbol{g}_{\mu\nu}^{\{\Psi\}}(\vec{x})$ and $\boldsymbol{\phi}_\Psi(\vec{x})$ act as linear operators. In terms of the classes of states that we have defined above, this can be turned into a sharp restriction: the *same* operators that represent the metric and other excitations in a state $|\Psi_g\rangle$ must also represent these excitations in superpositions (3.4) and (3.6). We show below that, in our construction, this is indeed the case.

To lighten the notation we now usually omit the superscript Ψ in $g_{\mu\nu}^{\{\Psi\}}$ even when we are considering state-dependent operators. Although in several cases we discuss explicitly whether a given operator is state dependent or state independent, in others it should be clear from the context.

We now point out that many of the existing methods of associating a geometry to a state as in (3.2) are state dependent in practice.[7] We hasten to add that this, by itself, does not mean that the map (3.2) can only be realized in a state-dependent manner. Our discussion in this subsection does *not* rule out the possibility that there may be an overarching state-independent prescription which encapsulates all of these state-dependent approaches in some approximation, or that the realizations of (3.2) that are discussed below cannot be interpreted as constructions which only hold in a limited class of states. Our purpose in this subsection is to use these examples to explain the distinction between state-dependent and state-independent realizations of the maps.

We now proceed to discuss the Ryu-Takayanagi formula, the procedure for extracting the Einstein equations from the first law of entanglement, and the smearing function construction of operators outside the black hole.

1. State-dependence in geometry from entanglement

The RT formula [24] and its generalization [27] by Hubeny, Rangamani and Takayanagi provide a method of reading off geometric quantities from a state. We review the formula, and show how it is state dependent. We also show how to interpret it correctly and that this state-dependence does not imply any contradiction with quantum mechanics.

In particular these formulas provide a relation between the entanglement entropy of a region on the boundary, and the area of an extremal surface in the bulk which is homologous to the boundary region. So, given a region R on the boundary and a semiclassical metric $g_{\mu\nu}$, we can calculate the area of this extremal area surface $A(g,R)$. The Ryu-Takayanagi formula now states

$$\frac{1}{4G_N}A(g,R) = S_R, \qquad \text{Ryu-Takayanagi} \tag{3.19}$$

where S_R is the entanglement entropy of the region R.

We now show the following.

(1) The formula (3.19) cannot be interpreted as an operator relation for the area, because there is no entanglement entropy operator.

(2) However, even though the entanglement entropy cannot, in general, be interpreted as the expectation value of a Hermitian operator, because of properties of the large-$\mathcal{N}$ CFT Hilbert space, we expect to find a state-dependent operator $\boldsymbol{A}_R$ in the CFT which has the property that

$$\langle\Psi|\boldsymbol{A}_R|\Psi\rangle = S_R(|\Psi\rangle),$$

both in states (3.2) and in superpositions of a small number of such states (3.4).

[6] We thank Nima Lashkari for this phrase.

[7] We cannot help making the curious observation that, within the string theory literature, this fact hardly attracted any attention or controversy until the recent discussions on the black hole interior.

We start by noting that if the metric is a state-independent operator, then the area of the minimal area surface, which is a functional of the metric, is also a state-independent operator. In fact, as we see below, from the point of view of a semiclassical quantization of gravity—which is what yields the justification for expecting the metric to be an ordinary operator—the area of the minimal area surface should be as good an operator as the metric. Therefore, we might expect the existence of some operator $\boldsymbol{A}_R$, so that in the state dual to the geometry with metric $g_{\mu\nu}$, we have

$$\boldsymbol{A}_R|\Psi_g\rangle = A(g,R)|\Psi_g\rangle + \mathrm{O}\left(\frac{1}{\mathcal{N}}\right).$$

However, on the other hand, the entanglement entropy is not a linear operator. The standard proof is as follows. Consider the division of the CFT Hilbert space into that of the region and its complement: $\mathcal{H} = \mathcal{H}_R \otimes \mathcal{H}_{\tilde{R}}$. Say that we want an operator $\boldsymbol{S}_R$ so that $\forall|\Psi\rangle \in \mathcal{H}$; we have $\langle\psi|\boldsymbol{S}_R|\Psi\rangle = S(|\Psi\rangle)$, where S is the entanglement entropy between R and its complement in that state. Now we note the following facts. Since $S(|\Psi\rangle)$ is always non-negative, the expectation value of the putative $\boldsymbol{S}_R$ operator is non-negative in all states; therefore it can have no negative eigenvalues. Second, we can find a complete basis of unentangled states

$$|\Psi_{ij}\rangle = |R_i\rangle \otimes |\tilde{R}_j\rangle, \tag{3.20}$$

where $i \in [1,\ldots \dim(\mathcal{H}_R)], j \in [1,\ldots \dim(\mathcal{H}_{\tilde{R}})]$. Clearly we expect $\langle\Psi_{ij}|\boldsymbol{S}_R|\Psi_{ij}\rangle = 0$. Moreover, since $|\Psi_{ij}\rangle$ is a basis, we also have $\mathrm{Tr}(\boldsymbol{S}_R) = 0$. Since $\boldsymbol{S}_R$ has no negative eigenvalues, and its trace is zero, it must be the case that $\boldsymbol{S}_R = 0$ identically. This is absurd. Therefore, there is no operator $\boldsymbol{S}_R$ whose expectation value equals the entanglement entropy in general. A simple extension of this argument shows that this is also true for the Renyi entropies $\mathrm{Tr}(\boldsymbol{\rho}_R^n)$, where $\boldsymbol{\rho}_R$ is the reduced density matrix of the region.

The fact that the entanglement entropy does not correspond to an ordinary linear operator may appear to be a formal statement, but it becomes acute in the following situation in the AdS/CFT correspondence. Consider a superposition of two different classical geometries, as in (3.4). For simplicity, we can consider a pure state which is a superposition of a pure state corresponding to a black hole at temperature β, with a corresponding metric g_β, and another pure state corresponding to a black hole at a temperature β', with a corresponding metric $g_{\beta'}$. Provided that $\beta - \beta' \gg \frac{1}{\mathcal{N}}$, we see that the corresponding pure states are almost orthogonal. We write the superposed state as

$$|\Psi_s\rangle = \alpha_1|\Psi_{g_\beta}\rangle + \alpha_2|\Psi_{g_{\beta'}}\rangle,$$

and normalizability requires $|\alpha_1|^2 + |\alpha_2|^2 = 1 + \mathrm{O}(e^{-\mathcal{N}})$. This is not a state that we usually consider, but it is certainly possible to consider such superpositions in the CFT since distinct geometries do not belong to strict superselection sectors.

From the bulk point of view, quantum mechanics provides the following prediction. If one measures the area in this state, one expects to find the answer $A(g_\beta, R)$ with probability $|\alpha_1|^2$ and $A(g_{\beta'}, R)$ with probability $|\alpha_2|^2$.

While the entanglement entropy cannot reproduce this probability distribution, with some work we can show that the entanglement entropy does correctly reproduce the expectation value of the area. The argument is as follows. Consider the reduced density matrix of the region R in all three states,

$$\begin{aligned}\boldsymbol{\rho}_R(\beta) &= \mathrm{Tr}_{\tilde{R}}(|\Psi_{g_\beta}\rangle\langle\Psi_{g_\beta}|),\\ \boldsymbol{\rho}_R(\beta') &= \mathrm{Tr}_{\tilde{R}}(|\Psi_{g_{\beta'}}\rangle\langle\Psi_{g_{\beta'}}|),\\ \boldsymbol{\rho}_R(\Psi_s) &= \mathrm{Tr}_{\tilde{R}}(|\Psi_s\rangle\langle\Psi_s|),\end{aligned}$$

where $\tilde{R}$ is the complement of R.

We can write both the states in terms of a Schmidt basis,

$$\begin{aligned}|\Psi_{g_\beta}\rangle &= \sum_i \gamma_i^\beta |R_i^\beta\rangle \otimes |\tilde{R}_i^\beta\rangle,\\ |\Psi_{g_{\beta'}}\rangle &= \sum_i \gamma_i^{\beta'} |R_i^{\beta'}\rangle \otimes |\tilde{R}_i^{\beta'}\rangle,\end{aligned} \tag{3.21}$$

where, by the definition of the Schmidt basis, we have

$$\begin{aligned}&\langle R_i^\beta|R_j^\beta\rangle = \delta_{ij}; \qquad \langle\tilde{R}_i^\beta|\tilde{R}_j^\beta\rangle = \delta_{ij};\\ &\langle R_i^{\beta'}|R_j^{\beta'}\rangle = \delta_{ij}; \qquad \langle\tilde{R}_i^{\beta'}|\tilde{R}_j^{\beta'}\rangle = \delta_{ij}\\ &\sum_i |\gamma_i^\beta|^2 = 1; \qquad \sum_i |\gamma_i^{\beta'}|^2 = 1.\end{aligned}$$

To simplify the analysis, without sacrificing anything of importance, let us truncate the range of i in (3.21) so that it runs over $\mathrm{O}(e^{\mathcal{N}})$ states. In almost any state, where the energy scales like $\mathcal{N}$, it is in fact true that even if the exact expansion of the state involves an infinite number of eigenvectors, all but an $\mathrm{O}(e^{\mathcal{N}})$ number of them are exponentially unimportant.

Now, the key point is that in a very large Hilbert space we expect that the Schmidt basis decomposition for the state $|\Psi_{g_\beta}\rangle$ and the state $|\Psi_{g_{\beta'}}\rangle$ is typically uncorrelated. This implies that

$$|\langle R_i^\beta|R_j^{\beta'}\rangle|^2 = \mathrm{O}(e^{-\mathcal{N}}); \qquad |\langle\tilde{R}_i^\beta|\tilde{R}_j^{\beta'}\rangle|^2 = \mathrm{O}(e^{-\mathcal{N}}). \tag{3.22}$$

Strictly speaking (3.22) is valid if one takes a large Hilbert space and divides it into two parts. In a local quantum field theory, it is possible that the very short distance modes in the two regions are entangled in a universal manner.

This does not affect any of our results since in considering the entanglement entropy we, in any case, must subtract off this universal part.

Now the first two reduced density matrices are given by

$$\boldsymbol{\rho}_R(\beta) = \sum_i |\gamma_i^\beta|^2 |R_i^\beta\rangle\langle R_i^\beta|,$$
$$\boldsymbol{\rho}_R(\beta') = \sum_i |\gamma_i^{\beta'}|^2 |R_i^{\beta'}\rangle\langle R_i^{\beta'}|.$$

The corresponding entanglement entropies are given by

$$S_\beta = -\mathrm{Tr}(\boldsymbol{\rho}_R(\beta)\ln\boldsymbol{\rho}_R(\beta)) = -2\sum_i |\gamma_i^\beta|^2 \ln|\gamma_i^\beta|,$$
$$S_{\beta'} = -\mathrm{Tr}(\boldsymbol{\rho}_R(\beta')\ln\boldsymbol{\rho}_R(\beta')) = -2\sum_i |\gamma_i^{\beta'}|^2 \ln|\gamma_i^{\beta'}|.$$

Moreover, we see that

$$\boldsymbol{\rho}_R(\Psi_s) = |\alpha_1|^2\boldsymbol{\rho}_R(\beta) + |\alpha_2|^2\boldsymbol{\rho}_R(\beta') + \boldsymbol{\rho}_{\mathrm{cross}},$$

where we see that the matrix involving the cross terms is

$$\boldsymbol{\rho}_{\mathrm{cross}} = \sum_{i,j}[\alpha_1\alpha_2^*\gamma_i^\beta(\gamma_j^{\beta'})^*\langle\tilde{R}_j^{\beta'}|\tilde{R}_i^{\beta'}\rangle]|R_i^\beta\rangle\langle R_j^{\beta'}| + \mathrm{H.c.}$$

Now even though this is an $e^{\mathcal{N}}\times e^{\mathcal{N}}$ sized matrix, we can check using (3.22) that $\mathrm{Tr}(\boldsymbol{\rho}_{\mathrm{cross}}) = \mathrm{O}(e^{-\mathcal{N}})$ and also that $\mathrm{Tr}(\boldsymbol{\rho}_{\mathrm{cross}}^2) = \mathrm{O}(e^{-\mathcal{N}})$. Therefore the cross terms have an exponentially small effect in the computations below, and we neglect them.

Now consider two positive integers m_1, m_2. We see that

$$\mathrm{Tr}(\boldsymbol{\rho}_R^{m_1}(\beta)\boldsymbol{\rho}_R^{m_2}(\beta')) = \sum_{i,j}|\gamma_i^\beta|^{2m_1}|\gamma_j^{\beta'}|^{2m_2}|\langle R_i^\beta|R_j^{\beta'}\rangle|^2.$$

Therefore, from (3.22), we see that

$$\mathrm{Tr}(\boldsymbol{\rho}_R^{m_1}(\beta)\boldsymbol{\rho}_R^{m_2}(\beta')) = \mathrm{O}(e^{-\mathcal{N}}), \quad \text{if } m_1, m_2 > 0.$$

This allows us to evaluate the entanglement entropy of the superposed state. In particular, using the result above, we see that mth Renyi entropy for the superposed state is given by

$$\mathrm{Tr}(\boldsymbol{\rho}_R(\Psi_s)^m) = |\alpha_1|^{2m}\mathrm{Tr}(\boldsymbol{\rho}_R(\beta)^m) + |\alpha_2|^{2m}\mathrm{Tr}(\boldsymbol{\rho}_R(\beta')^m) + \mathrm{O}(e^{-\mathcal{N}}).$$

Therefore the entanglement entropy is given by

$$\begin{aligned} S_R(\Psi_s) &= -\lim_{m\to1}\frac{d}{dm}\mathrm{Tr}(\boldsymbol{\rho}_R(\Psi_s)^m) \\ &= -|\alpha_1|^2\ln(|\alpha_1|^2)\mathrm{Tr}[\boldsymbol{\rho}_R(\beta)] - |\alpha_2|^2\ln(|\alpha_2|^2)\mathrm{Tr}[\boldsymbol{\rho}_R(\beta')] \\ &\quad - |\alpha_1|^2\mathrm{Tr}[\boldsymbol{\rho}_R(\beta)\ln(\boldsymbol{\rho}_R(\beta))] - |\alpha_2|^2\mathrm{Tr}[\boldsymbol{\rho}_R(\beta')\ln(\boldsymbol{\rho}_R(\beta'))] \\ &= -|\alpha_1|^2\ln(|\alpha_1|^2) - |\alpha_2|^2\ln(|\alpha_2|^2) + |\alpha_1|^2 S_R(\beta) + |\alpha_2|^2 S_R(\beta'). \end{aligned}$$

Therefore we see that

$$S_R(\Psi_s) = \frac{1}{4G_N}\langle A(R)\rangle - |\alpha_1|^2\ln(|\alpha_1|^2) - |\alpha_2|^2\ln(|\alpha_2|^2),$$

where $\langle A(R)\rangle = |\alpha_1|^2 A(g_\beta, R) + |\alpha_2|^2 A(g_{\beta'}, R)$ is the expectation value of the area obtained from a naive analysis.

In fact the additional term that we have obtained is always subleading even if we take a superposition of a large number of states. This is because the leading term is $\mathrm{O}(\mathcal{N})$ as we can see from the explicit factor of G_N in the formula above. Now, even if we superpose m-states in the form (3.4) with coefficients $\sum_{i=1}^m |\alpha_i|^2 = 1$, then the additional term is bounded by

$$-\sum_{i=1}^m |\alpha_i|^2\ln(|\alpha_i|^2) \le \ln(m).$$

Therefore, unless we take a superposition of an $e^{\mathcal{N}}$ number of states, we see that we can still consistently interpret the entanglement entropy as the expectation value of the operator that, classically, would correspond to the area,

$$S_R = \frac{1}{4G_N}\langle \boldsymbol{A}_R\rangle. \tag{3.23}$$

If we do take a superposition of an exponentially large number of states, then the cross terms become important even for the area operator, and we must reevaluate the entire expression.

To summarize, we have concluded that once the original Ryu-Takayanagi formula is interpreted as a relation between an expectation value and the entanglement entropy as in (3.23), then it holds consistently even in states that are superpositions of classical geometries as advertised. Our analysis here does not rule out the existence of a state-independent area operator $\boldsymbol{A}_R$ but such a state-independent operator cannot be dual to the entanglement entropy in general.

Before concluding, we should mention that there are several approaches that attempt to construct other bulk geometric quantities by massaging or refining the Ryu-Takayanagi formula. For example, the authors of [28] related the differential entropy—obtained by considering the variation of the entanglement entropy as the interval on the boundary is altered—to the area of a hole in the bulk. This can be used to read off the bulk metric more directly than the minimal area prescription. Of course, all of these approaches are also explicitly state dependent. However, just as in our discussion above, we expect that when we interpret them appropriately they do not present any observable contradiction with quantum mechanics in the bulk.

2. Equations of motion from the first law of entanglement

Another approach to deriving the bulk from the boundary, which has attracted attention, is the program of deriving the bulk equations of motion from the "first law of entanglement" [29–31]. Consider, once again, a region R on the boundary, and a CFT in the vacuum state. Then we may define the modular Hamiltonian of the region by demanding that the reduced density matrix of R has the form

$$\boldsymbol{\rho}_R = \frac{e^{-\boldsymbol{H}^R_{\rm mod}}}{\mathrm{Tr}_{\mathcal{H}_R}(e^{-\boldsymbol{H}^R_{\rm mod}})},$$

where the reader should note that the trace is in $\mathcal{H}_R$ only.

In this case, if we consider the vacuum of the CFT and take the region R to be a ball of radius a centered around a point $\vec{y}_0$, then the modular Hamiltonian is given by [32]

$$\boldsymbol{H}^R_{\rm mod} = 2\pi \int_R d^{d-1}\vec{y}\, \frac{a^2 - |\vec{y}-\vec{y}_0|^2}{2a}\boldsymbol{T}_{tt}, \tag{3.24}$$

where $\boldsymbol{T}_{tt}$ is the time-time component of the stress tensor. But this is a state-dependent formula that is obtained by defining the modular Hamiltonian about the vacuum.

Using this formula it was shown [29,33] that one can relate the linearized Einstein equations in the bulk to the first law of entanglement entropy under small changes of the state. By considering a generalization of the Ryu-Takayanagi conjecture, where the area is replaced by a Wald functional, this was extended to higher derivative theories in [30] and $1/\mathcal{N}$ interactions were included in [34].

However, although (3.24) looks like an operator equation, the modular Hamiltonian is also a state-dependent operator. There is no globally defined operator $\boldsymbol{H}^R_{\rm glob}$ in the theory so that its action equals that of the modular Hamiltonian on every possible state. The proof is similar to the one above. Let us say that we had an operator

$$\boldsymbol{H}^R_{\rm glob}|\Psi\rangle = \boldsymbol{H}^R_{\rm mod}|\Psi\rangle, \tag{3.25}$$

so that its action on $\mathcal{H}_R$ was that of the modular Hamiltonian and it acted as the identity on $\mathcal{H}_{\tilde{R}}$. Consider again the unentangled states in (3.20). The density matrix of R in this state is pure: $\boldsymbol{\rho}_R(|\Psi_{ij}\rangle) = |R_i\rangle\langle R_i|$. We can see that this implies that the putative modular Hamiltonian operator must have the action $\boldsymbol{H}^R_{\rm glob}|\Psi_{ij}\rangle = 0$. However, if $\boldsymbol{H}^R_{\rm glob}$ is a linear operator, then on any state $\boldsymbol{H}^R_{\rm glob}\sum_{ij}\alpha_{ij}|\Psi_{ij}\rangle = 0$. This suggests that $\boldsymbol{H}^R_{\rm glob} = 0$ as an operator, which is absurd. Therefore (3.25) cannot hold for any state-independent operator $\boldsymbol{H}^R_{\rm glob}$.

Therefore, (3.24) must be interpreted as a relation that is true within expectation values taken in the vacuum or small fluctuations about the CFT vacuum. No operator generalization of this equation exists as we have shown above. Nevertheless, it should be possible to obtain similar formulas about different states by defining the action of the modular Hamiltonian relative to that state. Such formulas also work for superpositions of a small number of states, as we showed above in the case of the entanglement entropy, but this entire process is fundamentally state dependent.

The authors of [35] proposed that $\boldsymbol{H}^R_{\rm mod}|\Psi\rangle = \boldsymbol{A}_R|\Psi\rangle$ should hold as an operator equation. However, as they noted explicitly this is a state-dependent relation which works in the neighborhood of a given state. As we discussed above we would also expect it to work in superpositions of a small number of semiclassical states.

3. Smearing function construction of local operators

Another commonly used method—and one that we use in this paper—of extracting local physics from a state uses a smearing function to represent bulk operators as smeared versions of boundary operators [36]. We review this approach in greater detail in Sec. IVB, where we also derive the expressions below for some states. In this approach, given a state $|\Psi_g\rangle$, we guess a smearing function and conjecture that local fields in the bulk have the form

$$\boldsymbol{\phi}(\vec{x}) = \int \boldsymbol{\mathcal{O}}(\vec{y}^b)K_g(\vec{y}^b,\vec{x})d^d\vec{y}^b, \tag{3.26}$$

where $\vec{x}$ is a bulk point, $\vec{y}^b$ is a boundary point, $\boldsymbol{\mathcal{O}}$ is a single-trace operator on the boundary, and K_g is an appropriately chosen smearing function. Strictly speaking, there are some difficulties in interpreting (3.26) in position space, having to do with the convergence of the integral, which has led to some confusion in the literature [37,38]. However, as we showed in [7], these difficulties go away if we work in momentum space and this subtlety is irrelevant for our present discussion.

One may object that one is putting in the answer by hand in (3.26) in the kernel K_g. However, it is a nontrivial fact that the operators $\boldsymbol{\phi}(\vec{x})$ do obey (3.10), and also have the right boundary values (as one approaches the boundary of

AdS) as CFT correlators. In particular for an operator $\mathcal{O}$ of dimension Δ, we require that

$$\langle \mathcal{O}(\vec{x}_1^b)\ldots\mathcal{O}(\vec{x}_n^b)\rangle = Z^n \lim_{r\to\infty} r_1^{\Delta}\ldots r_n^{\Delta}\langle \boldsymbol{\phi}(\vec{x}_1^b, r_1)\ldots\boldsymbol{\phi}(\vec{x}_n^b, r_n)\rangle, \tag{3.27}$$

where Z is a numerical wave-function renormalization factor, and we have written the bulk points as a boundary point combined with a radial coordinate r which can be identified with the coordinate r in (4.1). The fact that *both* (3.26)–(3.27) hold simultaneously involves a delicate interplay between the kernel and the correlators of $\mathcal{O}$ in the state $|\Psi_g\rangle$.

As written, the expression (3.26) is explicitly state dependent because the kernel K_g depends on the metric, and is therefore different in different states $|\Psi_g\rangle$. So, for a given kernel K_g, this expression works only in a state that corresponds to this semiclassical geometry.

In Sec. IV, we discuss whether it may be possible to lift (3.26) to a state-independent prescription, at least, outside the horizon. While this is possible in a minisuperspace approximation as we show around (4.21), we are agnostic about whether this works in general, even outside the horizon. In [3] it was argued that the $1/N$ corrections may automatically resum to give the correct smearing function on a general semiclassical background. It would be interesting to explore this possibility further and we comment more on this issue in [39].

C. A semiclassical obstruction to state-independence

Given that all existing examples of extracting local physics from the boundary involve various measurables, which are nevertheless not linear operators, why should we expect that the metric is given by an ordinary operator in the CFT? More precisely, what is the basis for the naive expectation that operators satisfying (3.8) and (3.10) should exist in the CFT? In this subsection, we try and explain the basis for this naive expectation, although, as we point out immediately, we believe that this intuition is flawed.

For simplicity, we consider whether one should expect a state-independent metric operator $\boldsymbol{g}_{\mu\nu}(\vec{x})$ to exist. A similar argument applies to other light fields in the theory.

The key point is that the classical metric $g_{\mu\nu}(\vec{x})$ is a well-defined function on the classical phase space of the theory. Recall that the classical phase space can be put in 1–1 correspondence with the set of all classical solutions of the theory. Given initial data for the canonical variables, and their conjugate momenta, we can evolve it forward to generate the entire classical solution. Conversely, given a classical solution, we can take a section by evaluating the variables and their momenta at some point in time to obtain a point on the phase space.

As we have explained above, once we go to a well-defined gauge, the value of the metric $g_{\mu\nu}(\vec{x})$ is well-defined in any classical geometry. Therefore the metric is a well-defined function on the phase space of the theory. Now, one usually expects that quantization takes functions on the phase space to well-defined operators in the Hilbert space. Therefore one might expect the metric $g_{\mu\nu}(\vec{x})$ in relational gauge to lift to a state-independent operator in the theory.

As we review in Appendix A, this is usually done as follows. In the quantum theory, we obtain coherent states, $|g\rangle$ corresponding to each semiclassical geometry. We then lift the classical function to an operator through

$$\boldsymbol{g}_{\mu\nu}(\vec{x}) \sim \sum_g g_{\mu\nu}(\vec{x})|g\rangle\langle g|, \tag{3.28}$$

where the sum is over all metrics, discretized in some fashion.[8]

Now, the analysis of Appendix A and Sec. VI shows that for such a construction to work, it is very important that if we consider the inner product of two distinct geometries, it dies off to arbitrarily small values

$$\langle g_1|g_2\rangle = e^{-\mathcal{N}\upsilon(g_1,g_2)}.$$

We can compute the function υ on the right-hand side in linearized gravity but in order for (3.28) to converge we require that for sufficiently distinct g_1, g_2, we can have $\upsilon \gg 1$.

On the other hand, in the CFT, as we have discussed coherent states of the metric $|g\rangle$ correspond to CFT states $|\Psi_g\rangle$. However, for generic states at the same energy $E \propto \mathcal{N}$, we have

$$\langle \Psi_{g_1}|\Psi_{g_2}\rangle = \mathrm{O}(e^{-\frac{S}{2}}),$$

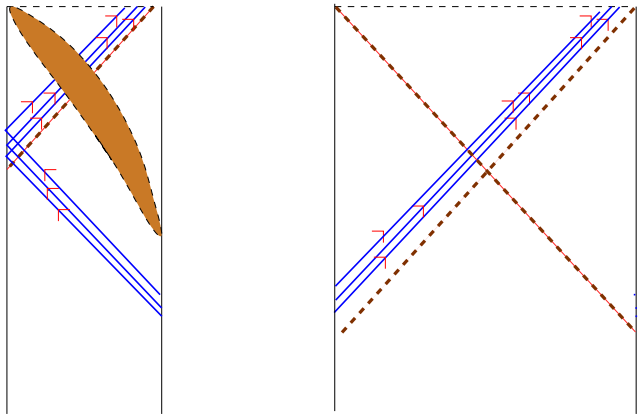
where $S \propto \mathcal{N}$ is the thermodynamic entropy of the CFT at the energy E.

This fat tail in the inner product of coherent states in the CFT subtly violates the expectation from a semiclassical quantization of gravity.[9] As a result of this tail, we cannot write down an expression of the form (3.28) with the putative coherent states replaced by $|\Psi_g\rangle$ because interference from distant microstates implies that the operator $\boldsymbol{g}_{\mu\nu}(\vec{x})$ on the left of (3.28) does not behave like the classical function $g_{\mu\nu}(\vec{x})$.

We direct the reader to Sec. VI for an example where this can be seen very clearly. In Appendix A we discuss the single-sided case in more detail and describe why we

[8]For a concrete example of a formula of this sort, the reader may wish to look at (4.21) although we caution the reader that (4.21) sums only over spherically symmetric metrics and works only outside the horizon. In contrast, we would like (3.28) to work for all kinds of metrics, and both inside and outside the horizon.

[9]This is reminiscent of the fact [40] that thermal correlators in the CFT decay down to $\mathrm{O}(e^{-S})$, in contrast to the naive expectation from semiclassical gravity that the exponential decay in time should continue forever.



(a) Hawking's derivation

(b) The analogy to the Eternal Black hole

FIG. 3. Two ways of arguing that new right movers are necessary behind a black hole horizon. Hawking's original derivation is on the left, where the right movers are modes that have "bounced" off $r = 0$ and propagated through the infalling matter. The analogy to the eternal black hole is on the right, where the right movers come from a left asymptotic region. Both of these suffer from difficulties, and so we perform a purely local derivation leading to the same result.

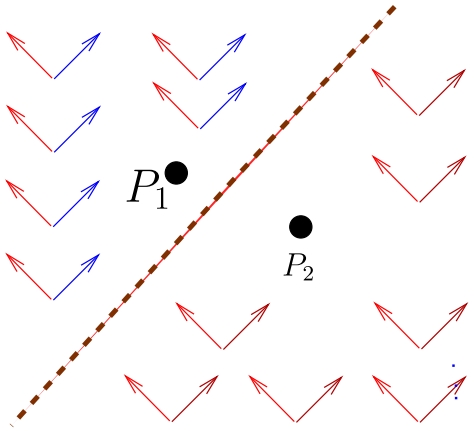


FIG. 4. We derive the necessity of new modes just by demanding a regular two-point function for points P_1, P_2 across the horizon without invoking another asymptotic region or tracing these modes back into the past.

believe that the same obstruction prevents one from writing down state-independent operators for well-defined classical geometric quantities.

IV. LOCAL BULK OPERATORS IN ADS/CFT: CONDITIONS FOR A SMOOTH INTERIOR

In this section, we review the conditions that are required to obtain a smooth exterior and interior geometry for a black hole in AdS/CFT. The central point that we emphasize in this section is that a smooth interior requires the existence of operators in the CFT, with specific properties that we enumerate below. We have dealt with this question in our previous papers [7–9], but we present a slightly new perspective here to buttress the same conclusion.

Before we proceed to the analysis, we briefly state our result and emphasize the difference with previous derivations. Consider a black hole horizon, which may have been formed due to gravitational collapse or may be part of an eternal black hole. If we quantize a field on both sides of the horizon, we find that while the Schwarzschild left movers cross the horizon smoothly, the Schwarzschild right movers do not. The claim is that to obtain a smooth horizon, we must find new operators, which play the role of right movers behind the horizon, and are appropriately entangled with the right movers in front of the horizon.

There are various ways to reach this conclusion. These right movers were identified in Hawking's original analysis of this question as modes from past null infinity that are concentrated in the time, just after the last null ray to escape the horizon. In Hawking's geometric analysis, these modes bounce from $r = 0$ to play the role of right movers behind the horizon. One can also argue for the existence of these right movers and the appropriate entanglement—as we did in [7]—by using the semiclassical intuition that, at late times, the collapsing geometry approaches the eternal black hole where these right movers originate from a left asymptotic region, which we call region III. Figure 3 displays the intuition for these arguments.

These derivations suffer from certain difficulties. Hawking's original work has a trans-Planckian problem because tracing these modes back to past null infinity boosts them to very high energies. Similarly, the intuition that these modes come from an effective region III is somewhat confusing because we do not expect any such region to exist for a collapsing geometry.

To solve these problems, in this section, we perform a purely local derivation that reveals the necessity of the existence of appropriate entangled right-moving modes behind the horizon. Our picture in this paper is shown in Fig. 4. We start with the sole assumption that the field in the near-horizon region outside and inside the horizon has an effective perturbative description. This assumption implies the universality of a certain two-point function. By Fourier transforming this universal two-point function, we infer that the right movers behind the horizon must exist, and also infer their two-point functions with modes in front of the horizon. We start by performing this analysis in the bulk, and then discuss the implications in the CFT.

A. Bulk analysis of the mirror operators

Let us start from the bulk perspective. We then examine how this must be translated to the boundary. For simplicity, let us consider a massless scalar field in the bulk. This analysis carries over, almost entirely unchanged to the case of the graviton and other fields.

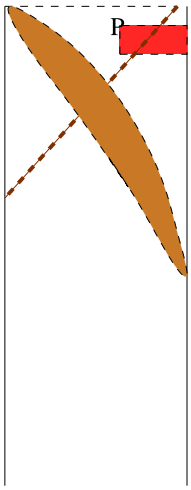


FIG. 5. We are interested in the late-time physics of the black hole geometry, schematically denoted by the rectangular patch P above.

Consider a big black hole in AdS. In the past this black hole could have been formed from the collapse of a star or some other physical process. However, we are interested in the late-time region shown schematically as the rectangular patch P in Fig. 5. This patch of spacetime overlaps with the region both in front of, and behind, the horizon. Classically, we expect that the initial collapsing matter, and any perturbations, have died away and are irrelevant for physics in this region. In the analysis below, we assume the validity of this classical expectation and derive various results for correlators of fields. Later we need to check the consistency of these results by ensuring that it is possible to construct a bulk to a boundary map that reproduces these correlators.

Geometry: The metric, at late times, outside the horizon is given by

$$ds^2 = -f(r)dt^2 + \frac{1}{f(r)}dr^2 + r^2 d\Omega_{d-1}^2, \tag{4.1}$$

where

$$f(r) = r^2 + 1 - c_d \frac{GM}{r^{d-2}},$$
$$c_d = \frac{8\pi^{\frac{2-d}{2}}\Gamma(d/2)}{d-1}.$$

The numerical constant, c_d, arises from the volume of the $d-1$-dimensional sphere. and we have set the radius of AdS to 1.

The horizon is defined implicitly by the equation $f(r_0) = 0$. As usual, it is convenient to introduce tortoise coordinates by $\frac{dr_*}{dr} = f^{-1}(r)$. Unlike in the case of the Schwarzschild black hole in flat space, we cannot usually express the tortoise coordinates in terms of the original coordinates using elementary functions. But we can choose the differential equation to satisfy

$$r_* = 0, \quad \text{at } r = \infty.$$

As $r \to r_0$, we see that $f^{-1}(r)$ diverges and $r_* \to -\infty$. In order to approach the future horizon we have to take the limit $r_* \to -\infty$ and at the same time $t \to +\infty$.

We introduce the following coordinates:

$$U = -e^{\frac{2\pi}{\beta}(r_* - t)}; \qquad V = e^{\frac{2\pi}{\beta}(r_* + t)}.$$

The horizon is given by $U = 0$, but with V being finite. We can check that with the factors of $\frac{2\pi}{\beta}$, the horizon is smooth in the U, V coordinate system. Near the horizon, with $(r - r_0) \ll 1$, we have $f(r) = \kappa(r - r_0)$. The constant κ is related to the temperature. A shortcut to determine the relation is to continue to Euclidean time, $t \to i\tau$, identify $\tau \sim \tau + \beta$ and make the change of variables $x = 2\sqrt{\frac{r-r_0}{\kappa}}$. Near the horizon, the analytically continued metric then takes the form

$$ds_E^2 \underset{x\to 0}{\to} dx^2 + \frac{\kappa^2}{4}x^2 d\tau^2 + r_0^2 d\Omega_{d-1}^2.$$

For the Euclidean circle τ to smoothly cap off at $x = 0$, we require $\frac{\kappa^2\beta^2}{4} = (2\pi)^2$ or $\kappa = \frac{4\pi}{\beta}$.

In the near-horizon region, we now have the following relations:

$$f(r) = \frac{4\pi}{\beta}(r - r_0), \qquad \Rightarrow r_* = \frac{\beta}{4\pi}\ln\left(\frac{r - r_0}{r_0}\right) + \text{const.}$$

From here, it follows that $f(r) \underset{r_*\to\infty}{\to} \kappa'(\frac{2\pi}{\beta})^2 e^{\frac{4\pi r_*}{\beta}}$, where κ' is another (irrelevant) constant.

In Kruskal coordinates the metric takes the form

$$ds^2 = \left(\frac{\beta}{2\pi}\right)^2 \frac{f(r)}{\text{UV}} dU dV + r^2 d\Omega_{d-1}^2,$$

and we see that the factor of $\frac{1}{\text{UV}}$ precisely cancels off the growing exponential in $f(r)$ near the horizon to ensure that the metric is regular.

$$g_{\mu\nu} \underset{U\to 0}{\to} -\kappa' dU dV + r_0^2 d\Omega_{d-1}^2.$$

After we cross the horizon, we can introduce a second Schwarzschild patch. Since $U > 0$ in the region inside the black hole (which we sometimes also call region II), we write

$$U = e^{\frac{2\pi}{\beta}(r_* - t)}; \qquad V = e^{\frac{2\pi}{\beta}(r_* + t)}, \qquad \text{in region II.}$$

Inside the horizon, the tortoise coordinate, r_*, rises from its value of $-\infty$ at the horizon, while the Schwarzschild time

decreases from its values of ∞ as one goes from right to left.

Two-point scalar correlators: Now, we consider a massless scalar field propagating in this background. We define this field using the relational prescription of Sec. III A 1. We derive various consequences of the fact that the horizon is smooth, simply by demanding that the two-point function both outside and inside the horizon be smooth.

We expect that the two-point scalar function has the form

$$\langle\phi(\vec{x}_1)\phi(\vec{x}_2)\rangle = G(\vec{x}_1,\vec{x}_2) + \mathrm{O}\left(\frac{1}{\mathcal{N}}\right).$$

We are interested in the regime where $\vec{x}_1$ and $\vec{x}_2$ approach the light cone, but always remain spacelike with respect to each other. In this regime the Wightman and time-ordered Green functions coincide and so we do not have to keep track of factors of $i\epsilon$. In the expression above, we have also used the fact that corrections to this expression come from interactions that are suppressed by $1/\mathcal{N}$. However, we do not need the full form of the propagator. For a large black hole, provided that the geodesic distance ℓ_{12} between $\vec{x}_1$ and $\vec{x}_2$ is small in comparison to the scale of curvature $\ell_{12} \ll \frac{1}{\beta}$, we expect that

$$\langle\phi(\vec{x}_1)\phi(\vec{x}_2)\rangle \approx \frac{1}{\left[g^{\mu\nu}(x_1-x_2)_\mu(x_1-x_2)_\nu\right]^{\frac{d-1}{2}}}, \quad |\ell_{12}| \ll \beta^{-1}. \tag{4.2}$$

Recall that the dimension of the bulk theory is $d+1$. The exponent above is the engineering dimension of the field, which is $\frac{(d+1)-2}{2}$. The relation (4.2) above is a powerful constraint, which holds in the short distance limit for any field theory in the bulk that is controlled by a free ultraviolet fixed point.[10]

Now we consider the correlation function as one point approaches the light cone of the other in the UV plane.[11] We work in the regime where the two points are separated on this plane so that $-(U_1-U_2)(V_1-V_2) > 0$.

$$\begin{aligned}&\langle\partial_{V_1}\phi(U_1,V_1,\Omega_1)\partial_{V_2}\phi(U_2,V_2,\Omega_2)\rangle\\ &= \partial_{V_1}\partial_{V_2}\frac{1}{\left(-\kappa'(U_1-U_2)(V_1-V_2)+\Omega_{12}^2\right)^{\frac{d-1}{2}}}\\ &= \frac{(d+1)(d-1)}{4}(\kappa')^2\frac{(U_1-U_2)^2}{\left(-\kappa'(U_1-U_2)(V_1-V_2)+\Omega_{12}^2\right)^{\frac{d+3}{2}}},\end{aligned}$$

where Ω_{12}^2 is defined as the distance between the points Ω_1 and Ω_2 on the sphere of radius r_0. We argue that this two-point function is actually proportional to a delta function in the coordinates on the sphere, as we take $U_1, U_2 \to 0$. If the transverse space had been planar, this would have been a planar delta function.

First note that we clearly have that

$$\lim_{U_1,U_2\to 0}\frac{(U_1-U_2)^2}{\left(-(U_1-U_2)(V_1-V_2)+\Omega_{12}^2\right)^{\frac{d+3}{2}}} = 0, \quad \text{for } \Omega_1 \neq \Omega_2.$$

But on the other hand, let us consider

$$I(U_1-U_2,V_1-V_2) = \int d^{d-1}\Omega_2\frac{(U_1-U_2)^2}{\left(-\kappa'(U_1-U_2)(V_1-V_2)+\Omega_{12}^2\right)^{\frac{d+3}{2}}}.$$

The integral above is on a sphere of radius r_0, but we can rescale the sphere by introducing a new variable $\Omega_2' = \frac{\Omega_2}{(\kappa'\delta)^{\frac{1}{2}}}$ with $\delta \equiv -(U_1-U_2)(V_1-V_2)$,

$$\begin{aligned}&I(U_1-U_2,V_1-V_2)\\ &= \int\left[\frac{(\kappa'\delta)^{\frac{d-1}{2}}(U_1-U_2)^2}{(\kappa'\delta)^{\frac{d+3}{2}}\left(1+\frac{\Omega_{12}^2}{\kappa'\delta}\right)^{\frac{d+3}{2}}}d^{d-1}\Omega_2'\right]\\ &= \frac{1}{(\kappa')^2(V_1-V_2)^2}\int\frac{d^{d-1}\Omega_2'}{\left(1+(\Omega_{12}')^2\right)^{\frac{d+3}{2}}}.\end{aligned}$$

The final integral is clearly a constant independent of Ω_1. This leads to the conclusion that

$$\begin{aligned}&\lim_{U_1-U_2\to 0}\langle\partial_{V_1}\phi(U_1,V_1,\Omega_1)\partial_{V_2}\phi(U_2,V_2,\Omega_2)\\ &= \kappa_N\frac{1}{(V_1-V_2)^2}\delta^{d-1}(\Omega_1-\Omega_2),\end{aligned}$$

where κ_N is a normalization constant that we do not fix here. In the same way, we also have

$$\begin{aligned}&\lim_{V_1-V_2\to 0}\langle\partial_{U_1}\phi(U_1,V_1,\Omega_1)\partial_{U_2}\phi(U_2,V_2,\Omega_2)\\ &= \kappa_N\frac{1}{(U_1-U_2)^2}\delta^{d-1}(\Omega_1-\Omega_2).\end{aligned} \tag{4.3}$$

This is a powerful and broadly applicable result. The ultralocality that we see in the transverse directions was also noted and used in the papers [41].

Now, let us see what this result implies for the correlation functions of the Schwarzschild creation and annihilation operators. Consider again the region near the horizon of a black hole, but this time in the original time and tortoise coordinates. Outside the horizon, we have the expansion

[10] Of course here we are talking about the intermediate regime, where $\ell_{12} \ll \beta^{-1}$ but at the same time $\ell_{12} \gg l_p, l_s$ where the latter are the Planck and string scales in the bulk.

[11] As we see below, to take this limit for correlators of the scalar itself is delicate, as a result of the usual complications of dealing with a massless scalar in two dimensions. This is the reason for taking correlators of its derivatives instead.

$$\phi(t,r_*,\Omega)\underset{U\to 0^-}{\to}\sum_m\int_0^\infty\frac{d\omega}{\sqrt{\omega}}a_{\omega,m}e^{-i\omega t}Y_m(\Omega)\times(e^{i\delta}e^{i\omega r_*}+e^{-i\delta}e^{-i\omega r_*})+\text{H.c.},\tag{4.4}$$

where $Y_m(\Omega)$ are spherical harmonics that we normalize below. The left and right movers get related to each other, and the phases δ depend on scattering in the black hole geometry [7]. As we noted above, and see again below, we can only use (4.4) for correlators of derivatives of the field.

Note that the canonical conjugate to the field outside the horizon is

$$\pi(t,r_*,\Omega)=g^{tt}\sqrt{-g}\frac{\partial}{\partial t}\phi(t,r_*,\Omega)=r^{d-1}\frac{\partial}{\partial t}\phi(t,r_*,\Omega).$$

We must impose the canonical commutation relations

$$\left[\phi(t,r_{*1},\Omega_1),\frac{\partial}{\partial t}\phi(t,r_{*2},\Omega_2)\right]=\frac{i}{r^{d-1}}\delta(r_{*1}-r_{*2})\delta^{d-1}(\Omega_1-\Omega_2).$$

Since the modes take this plane wave form in the near-horizon region, as $r\to r_0$, by imposing these commutation relations we find that they are satisfied only if

$$[a_{\omega,m},a^\dagger_{\omega',m'}]=\delta(\omega-\omega')\delta_{mm'},$$

provided that we normalize the spherical harmonics by

$$\sum_m Y_m(\Omega)Y_m^*(\Omega')=\frac{1}{4\pi r_0^{d-1}}\delta^{d-1}(\Omega-\Omega').$$

Now the two-point function, with both points outside the horizon but close to it, is given by

$$\langle\partial_{U_1}\phi(U_1,V_1,\Omega_1)\partial_{U_2}\phi(U_2,V_2,\Omega_2)\rangle=\frac{\beta^2}{4\pi^2U_1U_2}\sum_m\int_0^\infty\omega d\omega\times\left[(N_{\omega,m}+1)Y_m(\Omega_1)Y_m^*(\Omega_2)\left(\frac{U_1}{U_2}\right)^{\frac{i\beta\omega}{2\pi}}+N_{\omega,m}Y_m(\Omega_1)^*Y_m(\Omega_2)\left(\frac{U_2}{U_1}\right)^{\frac{i\beta\omega}{2\pi}}\right].\tag{4.5}$$

Here we have defined the two-point expectation value

$$\langle a^\dagger_{\omega,m}a_{\omega' m'}\rangle=N_{\omega,m}\delta(\omega-\omega')\delta_{m,m'}$$

in the black hole state and assumed that it is proportional to a delta function which is reasonable at late times when nothing depends on the time or the angular position.

Note that the expansion in two-point function (4.5) would not have converged without the derivatives on U_1, U_2. These derivatives pull down two factors of ω and ensure that the integrand is well behaved at $\omega=0$. Now we show that we must have

$$N_{\omega,m}=\frac{e^{-\beta\omega}}{1-e^{-\beta\omega}}.$$

To see this, note that

$$\int_0^\infty\omega d\omega\left(\frac{e^{-\beta\omega}}{1-e^{-\beta\omega}}\left(\frac{U_2}{U_1}\right)^{\frac{i\beta\omega}{2\pi}}+\frac{1}{1-e^{-\beta\omega}}\left(\frac{U_1}{U_2}\right)^{\frac{i\beta\omega}{2\pi}}\right)=\int_{-\infty}^\infty\omega d\omega\frac{e^{-\beta\omega}}{1-e^{-\beta\omega}}\left(\frac{U_2}{U_1}\right)^{\frac{i\beta\omega}{2\pi}}.$$

This integral can be completed in the lower half plane if $|U_1|>|U_2|$ and in the upper half plane otherwise. Picking up the poles at $\omega=\frac{2\pi i n}{\beta}$, we find that this integral evaluates to

$$\int_{-\infty}^\infty\omega d\omega\frac{e^{-\beta\omega}}{1-e^{-\beta\omega}}\left(\frac{U_2}{U_1}\right)^{\frac{i\beta\omega}{2\pi}}=-\frac{1}{\beta}\sum_n n\left(\frac{U_2}{U_1}\right)^n=-\frac{U_1U_2}{\beta(U_1-U_2)^2}.$$

Second, note that the sum over m in (4.5) automatically leads to a delta function proportional to $\delta^{d-1}(\Omega_1-\Omega_2)$. From the results above, we therefore find that (4.5) and (4.3) coincide provided that

$$\langle a_{\omega,m}a^\dagger_{\omega',m'}\rangle=\frac{1}{1-e^{-\beta\omega}}\delta(\omega-\omega')\delta_{mm'},\qquad\langle a^\dagger_{\omega,m}a_{\omega',m'}\rangle=\frac{e^{-\beta\omega}}{1-e^{-\beta\omega}}\delta(\omega-\omega')\delta_{mm'}.\tag{4.6}$$

Two caveats are in order. Note that (4.3) was derived in the near-horizon limit where $U_1,U_2\to 0$ and therefore our derivation above for the value of $N_{\omega,m}$ is not valid for low frequencies $\omega\ll\frac{1}{\beta}$. It is also not valid for Planckian frequencies $\omega=\mathrm{O}(\mathcal{N})$, where we do not expect effective field theory to give reliable results.

We now turn to the expansion behind the horizon. Here, as we quantize the field in region II, and approach the horizon from inside, we find an expansion.

$$\phi(t,r_*,\Omega)\underset{U\to 0^+}{\to}\sum_m\int_0^\infty\frac{d\omega}{\sqrt{\omega}}\times(a_{\omega,m}e^{-i\delta}e^{-i\omega(t+r_*)}Y_m(\Omega)+\tilde{a}_{\omega,m}e^{-i\delta}e^{i\omega(t-r_*)}Y_m^*(\Omega))+\text{H.c.}\tag{4.7}$$

Several points are worth noting in (4.7).

(1) By continuity of the mode $e^{i\omega(t+r_*)} = V^{\frac{i\beta\omega}{2\pi}}$, the operators a in region II must be the same as the operators in region I.
(2) Second, we need some operators to multiply the right-moving modes that vary as $e^{i\omega(t-r_*)}$. In (4.4) we identified these modes with $a_{\omega,m}$, but we find that this cannot be correct here. We call the $\tilde{a}_{\omega,m}$ operators the *mirror operators*.
(3) Note that the timelike coordinate inside the black hole is r_*. Therefore, the operator multiplying $e^{i\omega(t-r_*)}$ is classified as an "annihilation" operator. This is in spite of the fact that it has positive frequency with respect to t; the relevant point is that it has negative frequency with respect to r_*.
(4) Note that we have also conjugated the spherical harmonic Y_m for this mode. This is just a matter of choosing a convenient convention.

Inside the horizon, the canonical conjugate to the field is given by

$$\pi(t, r_*, \Omega) = g^{r_* r_*}\sqrt{-g}\frac{\partial}{\partial r_*}\phi(t, r_*, \Omega) = r^{d-1}\frac{\partial}{\partial r_*}\phi(t, r_*, \Omega).$$

The canonical commutation relations are

$$\left[\phi(t_1, r_*, \Omega_1), \frac{\partial}{\partial r_*}\phi(t_2, r_*, \Omega_2)\right] = \frac{i}{r^{d-1}}\delta(t_1 - t_2)\delta^{d-1}(\Omega_1 - \Omega_2).$$

By repeating the analysis of the canonical commutation relations we find that

$$[\tilde{a}_{\omega,m}, \tilde{a}^\dagger_{\omega',m'}] = \delta(\omega - \omega')\delta_{mm'},$$

where we have tacitly assumed that the possible mixed commutator $[\tilde{a}_{\omega,m}, a^\dagger_{\omega',m'}]$ vanishes. The mirror annihilation operator $\tilde{a}_{\omega,m}$ and the ordinary creation operator $a^\dagger_{\omega,m}$ have the same energy under the CFT Hamiltonian as we show in (4.13). So in a state that is time-translationally invariant, we do not expect this commutator to have a nonzero expectation value.[12]

We now consider a two-point function with one point in front of the horizon, and another point behind the horizon. This calls into play both the expansions (4.4) and (4.7). Recalling the fact that the relation between the Kruskal and Schwarzschild coordinates inside and outside the horizon differs by a minus sign, and repeating the derivation above for this case, we find that

$$\langle\partial_{U_1}\phi(U_1, V_1, \Omega_1)\partial_{U_2}\phi(U_2, V_2, \Omega_2)\rangle = \frac{\beta^2}{4\pi^2 U_1 U_2}\sum_{m,m'}\int_0^\infty \omega^{\frac{1}{2}}d\omega(\omega')^{\frac{1}{2}}d\omega'\mathcal{I}_{\omega,\omega',m,m'}, \quad (4.8)$$

with

$$\mathcal{I}_{\omega,\omega',m,m'} \equiv \langle a_{\omega,m}\tilde{a}_{\omega',m'}\rangle Y_m(\Omega_1)Y^*_{m'}(\Omega_2)(-U_1)^{\frac{i\beta\omega}{2\pi}}(U_2)^{\frac{-i\beta\omega'}{2\pi}} + \langle a_{\omega,m}\tilde{a}^\dagger_{\omega',m'}\rangle(-U_1)^{\frac{i\beta\omega}{2\pi}}(U_2)^{\frac{i\beta\omega'}{2\pi}}Y_m(\Omega_1)Y_{m'}(\Omega_2) + \text{H.c.} \quad (4.9)$$

Note that the result (4.3) is valid regardless of whether the points are on opposite sides, or the same side of the horizon. Now we find, repeating the contour integral argument above, that (4.8) agrees with (4.3) only if the two-point function between the two annihilation operators (and the two creation operators) is nonzero, whereas the mixed two-point function vanishes.

$$\langle a_{\omega,m}\tilde{a}_{\omega',m'}\rangle = \frac{e^{-\frac{\beta\omega}{2}}}{1 - e^{-\beta\omega}}\delta(\omega - \omega')\delta_{mm'}; \quad \langle a_{\omega,m}\tilde{a}^\dagger_{\omega',m'}\rangle = 0,$$
$$\langle a^\dagger_{\omega,m}\tilde{a}^\dagger_{\omega',m'}\rangle = \frac{e^{-\frac{\beta\omega}{2}}}{1 - e^{-\beta\omega}}\delta(\omega - \omega')\delta_{mm'}; \quad \langle a^\dagger_{\omega,m}\tilde{a}_{\omega',m'}\rangle = 0. \quad (4.10)$$

The additional factor of $e^{-\frac{\beta\omega}{2}}$ arises because of the relative minus sign between U_1 and U_2 in (4.9).

We can also consider the case where both points are inside the black hole. This is very similar to the cases above, so we just state the result. The smoothness of the two-point function of ϕ requires

$$\langle\tilde{a}_{\omega,m}\tilde{a}^\dagger_{\omega',m'}\rangle = \frac{1}{1 - e^{-\beta\omega}}\delta(\omega - \omega')\delta_{mm'},$$
$$\langle\tilde{a}^\dagger_{\omega,m}\tilde{a}_{\omega',m'}\rangle = \frac{e^{-\beta\omega}}{1 - e^{-\beta\omega}}\delta(\omega - \omega')\delta_{mm'}. \quad (4.11)$$

Finally, recall from the discussion of Sec. III A 1 that relationally defined observables in the bulk must obey the Heisenberg equations of motion. Consider a bulk point obtained considering a geodesic that originates on the boundary at point (t_b, Ω_b), with no initial velocity along the sphere, and following it for an affine parameter λ. In (3.17), this point was denoted by $P_\lambda(t_b, \Omega_b, \lambda)$. By solving the geodesic equation in the metric given by (4.1), we can trade these coordinates for Schwarzschild coordinates.

$$P_\lambda(t_b, \Omega_b, \lambda) = (t, \Omega, r_*).$$

Then it is easy to check that the isometry of the metric under time translations implies that if we follow another geodesic that originates at $t_b + T$, then

[12]This assumption of time-translational invariance on the boundary is not true in some cases, like in the geon geometry considered in [42] where the mirror operators can be identified with the ordinary ones.

$$P_\lambda(t_b+T,\Omega_b,\lambda)=(t+T,\Omega,r_*). \tag{4.12}$$

The relation (4.12) holds for points both outside and inside the horizon. In terms of the field this means that for the field written in Schwarzschild coordinates,

$$e^{iHT}\phi(t,r_*,\Omega)e^{-iHT}=\phi(t+T,r_*,\Omega),$$

where H is the boundary Hamiltonian that translates times on the boundary. This translates into the following commutation relations for the modes introduced above:

$$\begin{aligned}&[H,a_{\omega,m}]=-\omega a_{\omega,m};\qquad [H,a^\dagger_{\omega,m}]=\omega a^\dagger_{\omega,m},\\ &[H,\tilde{a}_{\omega,m}]=\omega\tilde{a}_{\omega,m};\qquad [H,\tilde{a}^\dagger_{\omega,m}]=-\omega\tilde{a}^\dagger_{\omega,m}.\end{aligned} \tag{4.13}$$

Note the opposite signs in the two lines of (4.13). This is a result of the fact that we mentioned above—the operator $\tilde{a}_{\omega,m}$ multiplies a mode that is positive frequency with respect to the Schwarzschild time.

Summary: In this section we considered a scalar field propagating in the geometry of a Schwarzschild black hole. By simply imposing the requirement that the two-point function had the correct short distance behavior we were able to derive necessary conditions on the two-point functions of the modes of the field in the black hole state. These conditions are given by (4.6) and (4.10)–(4.11). If the field is defined relationally with respect to the boundary, then the modes must also have the Hamiltonian commutators (4.13).

In the CFT we must find operators that satisfy these conditions in any state that is dual to a smooth geometry.

B. Local operators in the CFT

Let us now understand what the analysis above implies for the CFT. As discussed in Sec. III, we would like a family of operators in the CFT, parametrized by a set of real numbers, $\boldsymbol{\phi}(U,V,\Omega)$, so that the correlation functions of these operators reproduce the correlators of a perturbative field in AdS. In this subsection, we discuss how to find such correlators outside the horizon. We turn to the issue of the nature of these operators inside the horizon in Sec. V.

1. Local operators outside the horizon

For the CFT to successfully reproduce effective field theory correlators outside the horizon, it must have operators which play the role of the modes $a_{\omega,m}$ that we encountered in (4.4). If we allow ourselves to use state-dependent operators, then this can be done in a straightforward way, as we show below.

Dual to each propagating field in the bulk, we have a generalized free field (GFF), $\mathcal{O}$ on the boundary—usually it is a single trace operator in a gauge theory. The fact that bulk correlators factorize because the bulk theory is perturbative is reflected in the large-N factorization of boundary correlators. When evaluated in the vacuum,

$$\begin{aligned}&\langle 0|\mathcal{O}(t_1,\Omega_1)\ldots\mathcal{O}(t_{2n},\Omega_{2n})|0\rangle\\ &=\frac{1}{2^n}\sum_\pi\langle 0|\mathcal{O}(t_{\pi_1},\Omega_{\pi_1})\mathcal{O}(t_{\pi_2},\Omega_{\pi_2})|0\rangle\ldots\\ &\quad\times\langle 0|\mathcal{O}(t_{\pi_{2n-1}},\Omega_{\pi_{2n-1}})\mathcal{O}(t_{\pi_{2n}},\Omega_{\pi_{2n}})|0\rangle+\mathrm{O}\left(\frac{1}{\mathcal{N}}\right),\end{aligned} \tag{4.14}$$

where π sums over all possible permutations. A similar relation holds for thermal correlators.

$$\begin{aligned}&\frac{1}{Z(\beta)}\mathrm{Tr}[e^{-\beta H}\mathcal{O}(t_1,\Omega_1)\ldots\mathcal{O}(t_{2n},\Omega_{2n})]\\ &=\frac{1}{2^n}\sum_\pi\left(\frac{1}{Z(\beta)}\mathrm{Tr}[e^{-\beta H}\mathcal{O}(t_{\pi_1},\Omega_{\pi_1})\mathcal{O}(t_{\pi_2},\Omega_{\pi_2})]\ldots\right.\\ &\quad\left.\times\frac{1}{Z(\beta)}\mathrm{Tr}[e^{-\beta H}\mathcal{O}(t_{\pi_{2n-1}},\Omega_{\pi_{2n-1}})\mathcal{O}(t_{\pi_{2n}},\Omega_{\pi_{2n}})]\right)\\ &\quad+\mathrm{O}\left(\frac{1}{\mathcal{N}}\right).\end{aligned} \tag{4.15}$$

Note that (4.15) is subtly different from (4.14) and does not follow from it directly. In particular, in (4.15), the thermal two-point functions have *already* resummed the $\frac{1}{\mathcal{N}}$ series about the vacuum that appears in (4.14) into a *different* $\frac{1}{\mathcal{N}}$ series. In particular, the thermal two-point function

$$G_\beta(t_1,\Omega_1,t_2,\Omega_2)=\frac{1}{Z(\beta)}\mathrm{Tr}[e^{-\beta H}\mathcal{O}(t_1,\Omega_1)\mathcal{O}(t_2,\Omega_2)], \tag{4.16}$$

where $Z(\beta)$ is the partition function, is very different from the vacuum two-point function

$$G_{\mathrm{vac}}(t_1,\Omega,t_2,\Omega_2)=\langle 0|\mathcal{O}(t_1,\Omega_1)\mathcal{O}(t_2,\Omega_2)|0\rangle.$$

Also, note that the large-$\mathcal{N}$ factorization of the thermal correlators (4.15) may break down if the operators are separated by large distances in time.

Finally, by the usual equivalence of ensembles, and the eigenstate thermalization hypothesis [43], a similar statement holds when the thermal correlators on both sides of (4.15) are replaced by expectation values in a typical energy eigenstate of the CFT. Explicitly, this is the statement that in a typical eigenstate of the CFT $|E\rangle$ with energy $E\gg\mathcal{N}$, we again have

$$\begin{aligned}&\langle E|\mathcal{O}(t_1,\Omega_1)\ldots\mathcal{O}(t_{2n},\Omega_{2n})|E\rangle\\ &\quad=\frac{1}{Z(\beta)}\mathrm{Tr}[e^{-\beta H}\mathcal{O}(t_1,\Omega_1)\ldots\mathcal{O}(t_{2n},\Omega_{2n})]+\mathrm{O}\left(\frac{1}{\mathcal{N}}\right),\end{aligned}$$

where β is the temperature corresponding to the energy E. At high temperatures in the CFT we expect that this is given by

$$\beta = f_\beta\left(\frac{E}{\mathcal{N}}\right), \tag{4.17}$$

where f_β is a smooth function. For example, in the $\mathcal{N}=4$ super-Yang-Mills with $SU(N)$ gauge group at high temperature and at strong coupling on a sphere of volume V, we have

$$\beta = \left(\frac{8E}{3\pi^2 N^2 V}\right)^{-\frac{1}{4}}.$$

Therefore, in particular, correlators in an energy eigenstate also factorize, and the eigenstate two-point function is close to the thermal one. We use this important fact to switch freely between thermal and pure state expectations below.

Now consider the modes of these generalized free fields.

$$\mathcal{O}_{\omega_n,m} = \frac{1}{T_b^{\frac{1}{2}}}\int_{-T_b}^{T_b} \mathcal{O}(t,\Omega)e^{i\omega_n t}Y_m^*(\Omega)dtd^{d-1}\Omega. \tag{4.18}$$

Here we have discretized the modes by introducing a time band $[-T_b, T_b]$, and correspondingly we have introduced a discrete frequency $\omega_n = \frac{n}{T_b}$. This is necessary because if we consider the strict Fourier modes of the CFT operators, they do not have the behavior that we need below. In [8,9], we performed this discretization by "clubbing together" these Fourier modes, whereas here we have reverted to a time band that has some other advantages. We also need a UV cutoff on n because if we consider very high energy modes then the $\frac{1}{\mathcal{N}}$ corrections that we have neglected above become important.

Now we find that in eigenstates

$$\langle E|[\mathcal{O}_{\omega_n,m}, \mathcal{O}^\dagger_{\omega_{n'},m'}]|E\rangle = C_\beta(\omega_n,m)\delta_{\omega_n\omega_n'}\delta_{mm'} + \mathrm{O}(\mathcal{N}^{-1}).$$

On the right-hand side the delta functions follow from the fact that both sides have the same CFT energy and CFT angular momentum. The nontrivial coefficient $C_\beta(\omega_n, m)$ is a function of the temperature β corresponding to E by (4.17). Now we define the operators

$$\boldsymbol{a}_{\omega_n,m} = \frac{\mathcal{O}_{\omega_n,m}}{\sqrt{C_\beta(\omega_n,m)}} + \mathrm{O}(\mathcal{N}^{-1}). \tag{4.19}$$

These operators are the natural candidates for creation and annihilation operators in the bulk. By construction we have that up to $\mathcal{N}^{-1}$ corrections

$$[H, \boldsymbol{a}_{\omega_n,m}] = -\omega_n \boldsymbol{a}_{\omega_n,m}, \qquad [\boldsymbol{a}_{\omega_n,m}, \boldsymbol{a}^\dagger_{\omega_n',m'}] = \delta_{\omega_n,\omega_n'}\delta_{m,m'}.$$

It is not difficult to check that they have the right thermal two-point function.

$$\begin{aligned}\frac{1}{Z(\beta)}\mathrm{Tr}(e^{-\beta \boldsymbol{H}}\boldsymbol{a}_{\omega_n,m}\boldsymbol{a}^\dagger_{\omega_n,m}) &= \frac{1}{Z(\beta)}\mathrm{Tr}(\boldsymbol{a}^\dagger_{\omega_n,m}e^{-\beta \boldsymbol{H}}\boldsymbol{a}_{\omega_n,m}) \\ &= e^{\beta\omega_n}\frac{1}{Z(\beta)}\mathrm{Tr}(e^{-\beta \boldsymbol{H}}\boldsymbol{a}^\dagger_{\omega_n,m}\boldsymbol{a}_{\omega_n,m}) \\ &= e^{\beta\omega_n}\frac{1}{Z(\beta)}\mathrm{Tr}(e^{-\beta \boldsymbol{H}}\boldsymbol{a}_{\omega_n,m}\boldsymbol{a}^\dagger_{\omega_n,m}) \\ &\quad - e^{\beta\omega_n}\frac{1}{Z(\beta)}\mathrm{Tr}(e^{-\beta \boldsymbol{H}}),\end{aligned}$$

where we have used the cyclicity of the trace and the commutation relations above. A little algebra now shows that

$$\frac{1}{Z(\beta)}\mathrm{Tr}(e^{-\beta \boldsymbol{H}}\boldsymbol{a}_{\omega_n,m}\boldsymbol{a}^\dagger_{\omega_n,m}) = \langle E|\boldsymbol{a}_{\omega_n,m}\boldsymbol{a}^\dagger_{\omega_n,m}|E\rangle = \frac{1}{1-e^{-\beta\omega_n}},$$

where we have used the equivalence of ensembles and the relations above hold only up to $\frac{1}{\mathcal{N}}$ and other corrections from discretizations.

Now consider the CFT operator

$$\boldsymbol{\phi}(t, r_*, \Omega) = \sum_{\omega_n,m}\frac{1}{\sqrt{\omega_n}}\boldsymbol{a}_{\omega_n,m}f_{\omega_n,m}(t,r_*)Y_m(\Omega) + \mathrm{H.c.} \tag{4.20}$$

where $f_{\omega_n,m}$ is a solution of the Klein-Gordon equation in the metric (4.1) with the boundary condition at the horizon

$$f_{\omega_n,m} \underset{r\to r_0}{\to} (e^{i\delta}e^{i\omega_n r_*} + e^{-i\delta}e^{-i\omega_n r_*}),$$

and normalizable boundary conditions at infinity. The expansion (4.20) not only fulfils the necessary near-horizon conditions that we derived above; it also correctly reproduces the behavior of a bulk field propagating in a smooth spacetime in the rest of AdS. This completes our construction of local operators in a high energy eigenstate. As we mentioned in Sec. III B 3, we obtain a bonus, and a consistency check, from AdS/CFT. The fields constructed in (4.20), with the aid of (4.19), automatically satisfy

$$\begin{aligned}&\lim_{r\to\infty} r^{2\Delta}Z^2\langle E|\boldsymbol{\phi}(t_1,r_*,\Omega_1)\boldsymbol{\phi}(t_2,r_*,\Omega_2)|E\rangle \\ &\quad = W_\beta(t_1-t_2,\Omega_1,\Omega_2),\end{aligned}$$

where Z is a numerical factor and W_β is defined in (4.16). Note that we did not put this relation in by hand. It follows from, and is a prediction of, the claim that the eigenstate is dual to the black hole geometry.

2. A state-independent minisuperspace bulk-boundary map outside the horizon

In (4.19), we explicitly put in the commutator in the energy eigenstate. The modes in (4.20) also contain

information about the state. Therefore, as written the expression (4.20) is state dependent and will not correctly reproduce local correlation functions in states corresponding to black holes with macroscopically different properties.

Now we consider whether it is possible to write down an expansion that will work outside the horizon in a larger class of states. The basic idea is to use projectors to try and "detect" the state. We show how one can generalize (4.20) so that it works in all high energy spherically symmetric eigenstates.

Given a spherically symmetric energy eigenstate $|E\rangle$, we can associate a temperature to the energy eigenstate by means of (4.17), and also an associated metric via (4.1). We denote this metric as $g_{E,\mu\nu}$. We also consider modes $f_{E,\omega,m}$; these are the same as the modes $f_{\omega,m}$ in (4.20), except that we have displayed their energy dependence explicitly. Now, consider

$$\begin{aligned}\boldsymbol{\phi}_{\text{state-ind}}(t, r_*, \Omega) &= \sum_E \sum_{\omega,m} \frac{1}{\sqrt{\omega_n}} \left(\frac{1}{\sqrt{C_\beta(\omega_n, m)}} \mathcal{O}_{\omega_n,m} |E\rangle\langle E| \right) \\ &\times f_{E,\omega_n,m}(t, r_*) Y_m(\Omega) + \text{H.c.}, \end{aligned} \tag{4.21}$$

where, as we emphasized above, the expectation of the commutator that we have used to normalize the mode also depends on the energy eigenstate. The claim is that this generalizes the construction (4.20) so that, as long as we stay away from the horizon, it works in spherically symmetric states of the CFT corresponding to an arbitrary temperature.

To verify this, note that the expression (4.21) is designed so that when it acts directly on an energy eigenstate its action reduces to that of (4.20). Now consider an excitation of an energy eigenstate by a polynomial in the modes (4.18),

$$\mathcal{O}_{\omega_1,m_1} \ldots \mathcal{O}_{\omega_n,m_n} |E\rangle = \sum_i \alpha_i |E_i\rangle.$$

If $\sum n \ll \mathcal{N}$ and $\sum n\omega_n \ll \mathcal{N}$, then all states $|E_i\rangle$ that appear above have $\frac{E-E_i}{\mathcal{N}} = 0 + \mathrm{O}(\frac{1}{\mathcal{N}})$ and therefore, from (4.17), the coefficients α_i are restricted in support to states that have the same macroscopic temperature and correspond to the same macroscopic metric. Therefore, (4.21) again acts on this superposition as (4.20) away from the horizon. This is the expected behavior since we do not expect these excitations to have any significant backreaction on the geometry.

It is easy to verify that the action of (4.21) is also consistent with the fact that we expect states of the form (3.4) to behave like classical superpositions of different geometries.

If we approach too close to the horizon, then not all quantities of physical interest are smooth functions of the energy. For example, there has been some debate in the literature on highly spacelike modes [37] where the ratio of value of the mode function near the horizon to its value at the boundary can vary exponentially with temperature. Although we showed in [7] that these modes do not present an obstruction to reconstructing the field near the horizon in the thermal state, it is less clear how to deal with this difficulty in the putative state-independent expression (4.21). It is also not clear whether (4.21) can be refined to work in all nonspherically symmetric situations.

V. ARGUMENTS AGAINST STATE-INDEPENDENT OPERATORS

In the previous section we explicitly found operators $\boldsymbol{a}_{\omega_n,m}$ in the CFT that were dual to propagating modes in the bulk. However, if we want to describe local operators behind the horizon, then we also need to locate the operator $\tilde{\boldsymbol{a}}_{\omega_n,m}$ in the CFT. Alternately, we could find operators $\tilde{\mathcal{O}}_{\omega_n,m}$ related to $\tilde{\boldsymbol{a}}_{\omega_n,m}$ by a relation analogous to (4.19). At this order in $\frac{1}{\mathcal{N}}$, we do not have to consider corrections to (4.19) and we switch freely between $\tilde{\mathcal{O}}_{\omega,m}$ and $\tilde{\boldsymbol{a}}_{\omega,m}$.

In this section, we review and refine some of the arguments that suggest that these operators cannot be state independent in the CFT. In [2–4], these arguments were used to argue that the CFT could not look past the black hole horizon, or even more dramatically that the horizon was just a cloak for a "firewall." Our interpretation is, instead, that these arguments tell us that the bulk to boundary map is state dependent. From this point of view, the objective of this section is to prove that one must either accept state-dependence or firewalls.

A. Some general results regarding projectors

Before we continue with this analysis, let us remind the reader of some elementary properties about matrix elements of projection operators. Eigenvalues of projection operators are either 1 or 0, so the operator norm of a projection operator is $\|P\| = 1$. As a result projectors are bounded operators and this implies that the map from state vectors $|\Psi\rangle$ into expectation values $\langle\Psi|P|\Psi\rangle$ is a continuous map.

Hence, to the extent that we can characterize the physical properties of a state by evaluating expectation values of projectors, *nearby state vectors must have nearby physical properties*.

Let us try to make this a bit more precise. Suppose that we have two unit-normalized states $|\Psi_1\rangle$ and $|\Psi_2\rangle$ in the Hilbert space and we denote their difference as $|\delta\Psi\rangle = |\Psi_1\rangle - |\Psi_2\rangle$. We define $\delta = \||\delta\Psi\rangle\|$. We consider a projector and estimate the difference of its expectation value on the two nearby states,

$$\begin{aligned} &|\langle\Psi_1|P|\Psi_1\rangle - \langle\Psi_2|P|\Psi_2\rangle| \\ &\quad = |\langle\delta\Psi|P|\Psi_2\rangle + \langle\Psi_2|P|\delta\Psi\rangle + \langle\delta\Psi|P|\delta\Psi\rangle| \\ &\quad \le |\langle\delta\Psi|P|\Psi_2\rangle| + |\langle\Psi_2|P|\delta\Psi\rangle| + |\langle\delta\Psi|P|\delta\Psi\rangle| \\ &\quad \le 2\delta + \delta^2. \end{aligned}$$

Notice that it may also be useful to think of two nearby states as those obeying

$$|\langle\Psi_1|\Psi_2\rangle| = 1 - \frac{\epsilon^2}{2}, \tag{5.1}$$

with small positive ϵ. Since physical states are represented by rays on the Hilbert space, we are free to chose the phase of the vectors as we like. It is easy to check that there is a choice where $\epsilon = \delta$ and the same result as before follows i.e. for any two vectors obeying (5.1), we have

$$|\langle\Psi_1|P|\Psi_1\rangle - \langle\Psi_2|P|\Psi_2\rangle| \leq 2\epsilon + \epsilon^2. \tag{5.2}$$

We use these results below.

B. $N_a \neq 0$ argument

First, let us consider the $N_a \neq 0$ argument [4]. The essence of this argument is as follows. We would like the set of states in the CFT to obey two conditions, both of which seem motivated on physical grounds.

(1) Typical superpositions of energy eigenstates are not excited states from the point of view of the infalling observer.
(2) If we consider states that are eigenstates of a Schwarzschild number operator $\boldsymbol{N}_{\omega_n} \equiv \boldsymbol{a}^\dagger_{\omega_n,m}\boldsymbol{a}_{\omega_n,m}$, for the modes introduced in (4.18)–(4.19), then these are excited states from the point of view of the infalling observer.

To phrase the first condition more precisely consider the following set of energy eigenstates,

$$\mathcal{R}_E \equiv \{|E_i\rangle\colon\ E - \Delta \leq E_i \leq E + \Delta\},$$

where E is some mean energy and Δ is a spread. We use the same symbol $\mathcal{R}_E$ to denote the Hilbert space spanned by these states and the meaning should be clear from the context. We also denote

$$\mathcal{D}_E \equiv \dim(\mathcal{R}_E).$$

Finally we introduce

$$\boldsymbol{P}_E \equiv \text{projector onto}\, \mathcal{R}_E.$$

Now consider a projection operator $\boldsymbol{P}_F$ corresponding to the measurement of the infalling observer, defined so that $\boldsymbol{P}_F = 0$ corresponds to a smooth and empty interior. This projector can be constructed as an ordinary projector in the CFT Hilbert space if the bulk to boundary map is state independent. The authors of [4] used the number operator, as measured by the infalling observer, to detect whether the horizon was smooth but it is possible to use other operators and therefore we keep the analysis here general.

From the first physical assumption mentioned above, we expect that for typical states in $\mathcal{R}_E$ the expectation value of $\boldsymbol{P}_F$ should be small. Hence we expect

$$\frac{1}{\mathcal{D}_E}\text{Tr}_{\mathcal{R}_E}(\boldsymbol{P}_F) = 0 + \text{O}\left(\frac{1}{\mathcal{N}}\right). \tag{5.3}$$

The second condition means that for eigenstates $|N_i\rangle$ of the Schwarzschild number operator $\boldsymbol{N}_{\omega_n}$ we have

$$\langle N_i|\boldsymbol{P}_F|N_i\rangle = \text{O}(1). \tag{5.4}$$

In the large-$\mathcal{N}$ limit we have $[\boldsymbol{H}, \boldsymbol{N}_{\omega_n}] = 0 + O(\mathcal{N}^{-1})$, so we intuitively expect that we can find a basis of the Hilbert space $\mathcal{R}_E$ spanned by number operator eigenstates $|N_i\rangle$. The trace of an operator can be evaluated in any basis, so we can evaluate the trace (5.3) in the $|N_i\rangle$ basis. For each of the basis vectors (5.4) gives a significant contribution. Then it seems that we get

$$\frac{1}{\mathcal{D}_E}\text{Tr}_{\mathcal{R}_E}(\boldsymbol{P}_F) = \text{O}(1) + \text{small error} \tag{5.5}$$

and that hence typical states are not smooth, in contradiction to the first assumption above. This concludes the $N_a \neq 0$ argument of [4]. The result was interpreted by [4] as an indication that typical pure states do not have a smooth interior. The small error above is due to the fact that the operators $\boldsymbol{H}$ and $\boldsymbol{N}_{\omega_n}$ can be simultaneously diagonalized within $\mathcal{R}_E$ only in an approximate sense, in the large-N limit.

One might attempt to find a loophole in this argument by looking more carefully at the error terms mentioned above. Could it be that, contrary to what was assumed in [4], these error terms are significant enough to make the rhs of Eq. (5.5) close to zero? In the following subsection, we perform a systematic analysis of the error terms and exclude the possibility that they can invalidate the $N_a \neq 0$ argument.

1. Bounding errors in the $N_a \neq 0$ argument

The linear algebra literature contains several results on "almost commuting matrices" [44], which could be used to make the argument above rigorous. Here, rather than taking this path, we follow an approach motivated by perturbation theory to make the $N_a \neq 0$ paradox sharper.

We assume that

$$\boldsymbol{H} = \boldsymbol{H}_0 + \frac{1}{\mathcal{N}}V, \tag{5.6}$$

where the "infinite $\mathcal{N}$" Hamiltonian, $\boldsymbol{H}_0$, has the property that $[\boldsymbol{H}_0, \boldsymbol{N}_{\omega_n}] = 0$ and V is a "perturbation," whose matrix elements have the property that $\frac{\langle E|V|E\rangle}{E} = \text{O}(1)$ for nearby

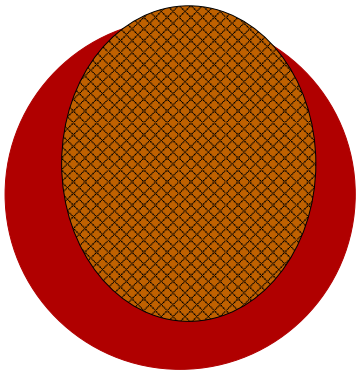


FIG. 6. The schematic structure of the two relevant sets. The solid circular set is the set of energy eigenstates. The smaller set of number eigenstates, shown as an elliptical patterned set, is almost completely contained inside the set of energy eigenstates.

high energy eigenstates of energy of order E.[13] Note that (5.6) is somewhat stronger than our original starting point—which was simply that $\langle E|[\boldsymbol{H},\boldsymbol{N}_{\omega_n}]|E\rangle = \mathrm{O}(\frac{1}{\mathcal{N}})$.[14]

If (5.6) is correct, then by standard arguments from perturbation theory we expect that groups of eigenstates of $\boldsymbol{H}$ can be reorganized into eigenstates of $\boldsymbol{N}_{\omega_n}$ and vice versa. Now consider the set of all number eigenstates that can be accurately approximated by energy eigenstates in $\mathcal{R}_E$. We call this set of $\boldsymbol{N}_{\omega_n}$ eigenstates $\mathcal{R}_-$ and denote its dimension by $\mathcal{D}_-$. The projector onto $\mathcal{R}_-$ is denoted by $\boldsymbol{P}_-$. By definition,

$$\langle N_i|\boldsymbol{P}_E|N_i\rangle = 1 - \mathrm{O}\left(\frac{1}{\mathcal{N}}\right), \quad \forall\ |N_i\rangle \in \mathcal{R}_-.$$

The structure of these two sets is shown in Fig. 6.

The key physical consequence of (5.6) is that to form eigenstates of $\boldsymbol{H}_0$ with an eigenvalue E, we have take eigenstates of $\boldsymbol{H}$ with $\boldsymbol{H}$-eigenvalues $E \pm \Delta$, where $\Delta = \mathrm{O}(\frac{E}{\mathcal{N}}) = \mathrm{O}(1)$. Therefore if we take the original spread of energies Δ in $\mathcal{R}_E$ to be large, $\Delta \gg \mathrm{O}(1)$, then we have

$$\frac{\mathcal{D}_E - \mathcal{D}_-}{\mathcal{D}_E} \ll 1. \tag{5.7}$$

If we accept these statements, then it is easy to produce a contradiction. From the assumptions above, given a $|N_i\rangle \in \mathcal{R}_-$, we have

$$|N_i\rangle = \sum_m U^*_{mi}|E_m\rangle = \sum_{m\in\mathcal{R}_E} U^*_{mi}|E_m\rangle + \sum_{m\notin\mathcal{R}_E} U^*_{mi}|E_m\rangle \equiv |M_i\rangle + |R_i\rangle,$$

where U^*_{mi} is some matrix that implements the change in the two eigenvalue bases and where $\langle R_i|R_i\rangle = \mathrm{O}(\frac{1}{\mathcal{N}})$. Here, we have divided the sum into two parts and used the definition of $\mathcal{R}_-$ which is precisely that its elements can be reexpressed as elements in $\mathcal{R}_E$. Moreover, using (5.2) we find that $\langle M_i|\boldsymbol{P}_F|M_i\rangle = \langle N_i|\boldsymbol{P}_F|N_i\rangle + \mathrm{O}(\frac{1}{\mathcal{N}})$. But this implies that

$$\frac{1}{\mathcal{D}_E}\mathrm{Tr}(\boldsymbol{P}_-\boldsymbol{P}_E\boldsymbol{P}_F\boldsymbol{P}_E\boldsymbol{P}_-) = \frac{1}{\mathcal{D}_E}\mathrm{Tr}(\boldsymbol{P}_-\boldsymbol{P}_F\boldsymbol{P}_-) = \frac{1}{\mathcal{D}_E}\mathrm{Tr}(\boldsymbol{P}_F\boldsymbol{P}_-) = \kappa\frac{\mathcal{D}_-}{\mathcal{D}_E},$$

where κ is some constant of $\mathrm{O}(1)$ which determines the probability for an infalling observer to see an excitation in a number eigenstate and which follows from (5.4). Here we have neglected $\mathrm{O}(\frac{1}{\mathcal{N}})$ corrections.[15]

Second, notice that the original trace in the microcanonical ensemble can be transformed by a sequence of elementary manipulations to

$$\begin{aligned}\mathrm{Tr}(\boldsymbol{P}_F\boldsymbol{P}_E) &= \mathrm{Tr}(\boldsymbol{P}_E\boldsymbol{P}_F\boldsymbol{P}_E) = \mathrm{Tr}((1-\boldsymbol{P}_- + \boldsymbol{P}_-)\boldsymbol{P}_E\boldsymbol{P}_F\boldsymbol{P}_E)\\ &= \mathrm{Tr}((1-\boldsymbol{P}_-)\boldsymbol{P}_E\boldsymbol{P}_F\boldsymbol{P}_E) + \mathrm{Tr}(\boldsymbol{P}_-\boldsymbol{P}_E\boldsymbol{P}_F\boldsymbol{P}_E)\\ &= \mathrm{Tr}((1-\boldsymbol{P}_-)\boldsymbol{P}_E\boldsymbol{P}_F\boldsymbol{P}_E(1-\boldsymbol{P}_-))\\ &\quad + \mathrm{Tr}(\boldsymbol{P}_-\boldsymbol{P}_E\boldsymbol{P}_F\boldsymbol{P}_E\boldsymbol{P}_-).\end{aligned}$$

Here we have repeatedly used the cyclicity of the trace, and the fact that projectors square to themselves. Now notice that given any product of projectors $X = P_1 \ldots P_n$, we find that $\mathrm{Tr}(X) = \mathrm{Tr}(X^\dagger X) \geq 0$. Therefore the first term in the last line above is positive and we find

$$\mathrm{Tr}(\boldsymbol{P}_F\boldsymbol{P}_E) \geq \mathrm{Tr}(\boldsymbol{P}_-\boldsymbol{P}_E\boldsymbol{P}_F\boldsymbol{P}_E\boldsymbol{P}_-) = \kappa\frac{\mathcal{D}_-}{\mathcal{D}_E}. \tag{5.8}$$

Combing the result of (5.8) and the physical assumption (5.3), we seem to find

[13]Note that if we wish to ensure that we can carry out perturbation theory to higher orders, we would also like V to obey the eigenstate thermalization hypothesis described in greater detail in (7.15).

[14]It is subtle to consider perturbations of the Hilbert space at high energies in $\frac{1}{\mathcal{N}}$ because the Hilbert space changes discontinuously with $\mathcal{N}$ and its dimension goes off to ∞ as $\mathcal{N} \to \infty$. So we are assuming that (5.6) holds at each $\mathcal{N}$ and some properties of these operators, such as the *ratio* of the dimensions of different sets below have a well-defined large-$\mathcal{N}$ limit.

[15]In the equation above, the first equality can be understood as follows: we have $\frac{1}{\mathcal{D}_E}\mathrm{Tr}(\boldsymbol{P}_-\boldsymbol{P}_E\boldsymbol{P}_F\boldsymbol{P}_E\boldsymbol{P}_-) = \frac{1}{\mathcal{D}_E}\sum_{i\in\mathcal{R}_-}\langle N_i|\boldsymbol{P}_E\boldsymbol{P}_F\boldsymbol{P}_E|N_i\rangle = \frac{1}{\mathcal{D}_E}\sum_{i\in\mathcal{R}_-}\langle M_i|\boldsymbol{P}_F|M_i\rangle$ and on the other hand we have $\frac{1}{\mathcal{D}_E}\mathrm{Tr}(\boldsymbol{P}_-\boldsymbol{P}_F\boldsymbol{P}_-) = \frac{1}{\mathcal{D}_E}\sum_{i\in\mathcal{R}_-}\langle N_i|\boldsymbol{P}_F|N_i\rangle = \frac{1}{\mathcal{D}_E}\sum_{i\in\mathcal{R}_-}\langle M_i|\boldsymbol{P}_F|M_i\rangle + \frac{1}{\mathcal{D}_E}\sum_{i\in\mathcal{R}_-}2\mathrm{Re}[\langle R_i|\boldsymbol{P}_F|M_i\rangle] + \frac{1}{\mathcal{D}_E}\sum_{i\in\mathcal{R}_-}\langle R_i|\boldsymbol{P}_F|R_i\rangle]$. Now, $\boldsymbol{P}_F$ is a projector operator so from the discussion of Sec. V A and the fact that the norm of the state $|R_i\rangle$ is $\mathrm{O}(\frac{1}{\mathcal{N}})$ we learn that each of the terms in the sums over i is small. Finally, the number of terms in this sum is $\mathcal{D}_-$ so from (5.7) we learn that the last two sums on the rhs are unimportant, thus establishing the desired result.

$$0 = \mathrm{Tr}(\boldsymbol{P}_F \boldsymbol{P}_E) \geq \kappa \frac{\mathcal{D}_-}{\mathcal{D}_E}. \tag{5.9}$$

This is clearly a contradiction, if we recall (5.7). Note that the difference between the left and right sides of (5.9) is O(1), and so the errors, which we have bounded to be $\mathrm{O}(\frac{1}{\mathcal{N}})$ using the construction above cannot affect this result.

This was used by [4] to suggest that (5.3) should be abandoned. We show below how a more plausible explanation is that $\boldsymbol{P}_F$ does not exist as a fixed (state-independent) linear projector; rather the question of whether a firewall exists or not depends on a state-dependent measurable.

C. Negative occupancy argument

We now present an argument that is closely related to the counting argument (or the lack of a left-inverse argument). As originally stated in [3], the counting argument is as follows. First, we consider a mode behind the horizon with creation and annihilation operators obeying the algebra

$$[\tilde{\boldsymbol{a}}_{\omega_n,m}, \tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}] = 1. \tag{5.10}$$

Notice that this equation unambiguously selects $\tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}$ as the creation operator, since we can rewrite it as $[(1 + \tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}\tilde{\boldsymbol{a}}_{\omega_n,m})^{-1}\tilde{\boldsymbol{a}}_{\omega_n,m}]\tilde{\boldsymbol{a}}^\dagger_{\omega_n,m} = 1$, which means that the operator $\tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}$ has a left inverse and hence it does not annihilate any state.

Then we notice that, as explained in Sec. IV, modes behind the horizon obey inverted commutators with the CFT Hamiltonian

$$[\boldsymbol{H}, \tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}] = -\omega_n \tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}. \tag{5.11}$$

This means that the operator $\tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}$, despite being a creation operator, lowers the energy of the CFT. Hence, it maps the space of states of energy E into that of energy $E - \omega_n$. However, the density of states in the CFT increases monotonically with energy. This implies that the operator $\tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}$ maps the larger Hilbert space of energy E into a smaller one of energy $E - \omega_n$. The *linear* operator $\tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}$ can do this only if it annihilates a fraction of the states of energy E. But this is in contradiction with the prediction of (5.10) that $\tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}$ has a left inverse.

Hence it seems that imposing the algebra (5.10)–(5.11) for *state-independent* linear operators is inconsistent with the growth of entropy in the CFT. This concludes the counting argument of [3].

One apparent difficulty with this argument is that it is phrased in terms of operator relations (5.10)–(5.11). One might wonder whether it is possible to satisfy these relations, not as operator equations, but only *within simple correlation functions*. We now present a closely related argument that is phrased entirely within the context of low point correlation functions.

Let $\boldsymbol{P}_E$ be the projector onto a narrow band of energy states. Define $\mathcal{D}_E = \mathrm{Tr}(\boldsymbol{P}_E)$, which counts the number of states in this band. We consider the expectation value of the occupation level of the mode in this ensemble of states,

$$\begin{aligned}\langle \tilde{\boldsymbol{N}}_{\omega_n} \rangle &= \mathcal{D}_E^{-1}\mathrm{Tr}(\boldsymbol{P}_E \tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}\tilde{\boldsymbol{a}}_{\omega_n,m}) = \mathcal{D}_E^{-1}\mathrm{Tr}(\tilde{\boldsymbol{a}}_{\omega_n,m}\boldsymbol{P}_E\tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}) \\ &= \mathcal{D}_E^{-1}\mathrm{Tr}(\boldsymbol{P}_{E+\omega n}\tilde{\boldsymbol{a}}_{\omega_n,m}\tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}) + \delta_1 \\ &= e^{\beta\omega_n} + \mathcal{D}_E^{-1}\mathrm{Tr}(\boldsymbol{P}_{E+\omega n}\tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}\tilde{\boldsymbol{a}}_{\omega_n,m}) + \delta_1 + \delta_2.\end{aligned} \tag{5.12}$$

In the first line we used the cyclicity of the trace. In the second line we used that (5.11) should hold inside simple correlators, which implies $\tilde{\boldsymbol{a}}_{\omega_n,m}\boldsymbol{P}_E = \boldsymbol{P}_{E+\omega n}\tilde{\boldsymbol{a}}_{\omega_n,m}$ up to some small error δ_1. In the last line we used that (5.10) should hold in simple correlators, up to some small error δ_2. Since the trace above consists just of a sum of low point correlators we expect that $\delta_1, \delta_2 \sim \mathrm{O}(\frac{1}{\mathcal{N}})$. This assumption allows us to ignore these errors in deriving the contradiction that follows. The factor outside the trace of $e^{\beta\omega_n}$ arises because

$$\mathcal{D}_E^{-1}\mathrm{Tr}(\boldsymbol{P}_{E+\omega n}) = \frac{\mathcal{D}_{E+\omega n}}{\mathcal{D}_E} = e^{\beta\omega_n}.$$

We also use the fact that for a reasonably smooth operator $\tilde{\boldsymbol{N}}_{\omega_n}$, we have

$$\mathcal{D}_E^{-1}\mathrm{Tr}(\boldsymbol{P}_{E+\omega_n}\tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}\tilde{\boldsymbol{a}}_{\omega_n,m}) = e^{\beta\omega_n}\langle \tilde{\boldsymbol{N}}_{\omega_n} \rangle + \mathrm{O}(\mathcal{N}^{-1}).$$

Replacing this in (5.12) and dropping all subleading error terms we arrive at our final relation

$$\langle \tilde{\boldsymbol{N}}_{\omega_n} \rangle = e^{\beta\omega_n} + e^{\beta\omega_n}\langle \tilde{\boldsymbol{N}}_{\omega_n} \rangle \Rightarrow \langle \tilde{\boldsymbol{N}}_{\omega_n} \rangle = -\frac{1}{1 - e^{-\beta\omega_n}},$$

which is negative. In some sense, this unphysical result is not surprising, because $\tilde{\boldsymbol{a}}_{\omega_n,m}$ is an annihilation operator with positive energy, and the thermal properties of such an operator seem to be ill defined.

To summarize, the argument above demonstrates that there cannot exist *linear, state-independent* operators in the CFT which approximately satisfy the relations (5.10)–(5.11) inside simple correlation functions. One might conclude from this that the black hole does not have an interior that the CFT can describe. Instead, we advocate [7–9] that the desired relations (5.10)–(5.11) can be consistently realized by allowing the operators $\tilde{\boldsymbol{a}}_{\omega_n,m}, \tilde{\boldsymbol{a}}^\dagger_{\omega_n,m}$ to depend on the state. For state-dependent operators the counting argument does not apply [9] and the negative occupancy argument presented above does not apply since it is meaningless to evaluate the trace, if the operators vary as a function of the state in the ensemble.

D. The generic commutator

Now we consider the fact that there is not enough space in the CFT Hilbert space to accommodate the commutant of the ordinary operators if they are finely spaced enough. There are two ways in which this argument can be phrased. One point, which was originally made in [3] is as follows. If we assume that the algebra of the mirror operators is given by some "scrambling" unitary transform of the ordinary operators so that we have

$$\tilde{\boldsymbol{a}}^{\dagger}_{\omega_n,m} = \boldsymbol{U}\boldsymbol{a}^{\dagger}_{\omega_n,m}\boldsymbol{U}^{\dagger},$$

then we find that, for a *generic* unitary operator $\boldsymbol{U}$, we have

$$|[\tilde{\boldsymbol{a}}^{\dagger}_{\omega_n,m}, \boldsymbol{a}_{\omega_n,m}]|^2 \sim \mathrm{O}(1).$$

This by itself is not a proof of the lack of existence of the commutant. In particular, if the Hilbert space has a factorization into coarse and fine pieces, as was discussed originally in [7], then this would break down.

In what follows, we discuss how finely an observer has to measure generalized free fields on the boundary, in order to exhaust the space of the CFT. However, first, we turn to two toy models: the spin chain and a set of decoupled harmonic oscillators.

Consider a chain of spins. We denote the operators acting on this chain by σ^i_a as in [9]. We assume that the spins are all decoupled. The index $i = 1\ldots N$, where N is the length of the spin chain, and $a = x, y, z$ as usual. We normalize them to satisfy $[\sigma^i_a, \sigma^j_b] = \frac{i}{2}\delta^{ij}\epsilon_{abc}\sigma^i_c$. A complete set of operators for the Hilbert space is obtained by taking arbitrary products of these single-spin operators. Nevertheless, even if we consider the significantly smaller set of just the N single-spin operators, the commutant of this smaller set is trivial and consists only of the identity operator.

One might hope that there exist (state-independent) operators $\tilde{\sigma}$, apart from the identity, which *approximately* commute with all single-spin operators. We now demonstrate that this is not possible: if $\tilde{\sigma}$ has small commutators with all single-spin operators, then $\tilde{\sigma}$ is small as an operator. To show this, we consider an arbitrary operator $\tilde{\sigma}$ acting on the spin chain. In order to factor out the identity operator, which is trivially in the commutant, we assume that $\tilde{\sigma}$ is traceless, which means that we can represent it as a polynomial in the atomic spin operators

$$\tilde{\sigma} = \sum_{i_m,a_m,n} c^{a_1\ldots a_n}_{i_1\ldots i_n}\sigma^{i_1\ldots i_n}_{a_1\ldots a_n},$$

where $\sigma^{i_1\ldots i_n}_{a_1\ldots a_n} \equiv \sigma^{i_1}_{a_1}\ldots\sigma^{i_n}_{a_n}$, and we impose the constraint that $i_1 < i_2 < \ldots i_n$ to avoid overcounting.

We find that we have the following relation:

$$[\tilde{\sigma}, \sigma^j_b] = \frac{i}{2}\sum c^{a_1\ldots a_n}_{i_1\ldots i_n} \times (\delta^j_{i_1}\epsilon_{a_1bc}\sigma^{i_1}_c\sigma^{i_2\ldots i_n}_{a_2\ldots a_n} + \delta^j_{i_2}\epsilon_{a_2bc}\sigma^{i_2}_c\sigma^{i_1 i_3\ldots i_n}_{a_1 a_3\ldots a_n} + \cdots).$$

While we have written a sum of delta functions on the right, note that at most one of them is nonvanishing. A natural norm of an operator to consider in this space is $|X|^2 = \frac{1}{2^n}\mathrm{Tr}(X^{\dagger}X)$. With this definition

$$|[\tilde{\sigma}, \sigma^j_b]|^2 = \frac{1}{4}\sum |c^{a_1\ldots a_n}_{i_1\ldots i_n}\delta^j_{i_1}\epsilon_{a_1bc}|^2 + |c^{a_1\ldots a_n}_{i_1\ldots i_n}\delta^j_{i_2}\epsilon_{a_2bc}|^2 + \cdots$$

Note that there is no interference between the different terms in the sum due to the observation above. However, when we sum over b we find that there are two values for which the completely antisymmetric tensor is nonzero. This leads to

$$\sum_{j,b}|[\tilde{\sigma}, \sigma^j_b]|^2 = \frac{1}{2}\sum |c^{a_1\ldots a_n}_{i_1\ldots i_n}|^2 = \frac{1}{2}|\tilde{\sigma}|^2.$$

The physical implication of this is as follows. If an observer can measure the various single-spin operators, then given any operator $\tilde{\sigma}$, the observer can detect that it fails to commute with these ordinary operators. In particular, it is *not necessary* for the observer to measure very complicated observables. Even if the observer does not have access to more complicated products of these spin operators, she can determine that the commutant is trivial.

The argument presented above shows that an operator of unit norm, $|\tilde{\sigma}|^2$, must have an order 1 commutator with at least one single-spin operator, or alternatively it could have $\mathrm{O}(\frac{1}{S})$ commutators with all the single-spin operators. In either case, the important point is that it cannot simultaneously have smaller commutators with all the σ^i_a.

Now, we consider a similar argument for the case of decoupled harmonic oscillators. The setup was described in more detail in [9]. We have unbounded creation and annihilation operators. The frequencies of the oscillators are given by $\omega_1\ldots\omega_N$ and their respective creation and annihilation operators are specified by $a_1\ldots a_N$. The only nonzero commutators are $[a_i, a^{\dagger}_j] = \delta_{ij}$. The Hilbert space is a Fock space indexed by the eigenvalues of the number operators $N_i = a^{\dagger}_i a_i$.

We can still write any operator of interest as

$$\tilde{a} = \sum_{p_j,q_j} A(p_1, q_1\ldots p_n, q_n)a_1^{p_1}(a_1^{\dagger})^{q_1}\ldots a_N^{p_N}(a_N^{\dagger})^{q_N}.$$

Once again we factor out factors of N_i from each monomial in the polynomial above so that either $p_i = 0$ or $q_i = 0$ for all i. the most general operator then

lives in the direct product of the vector space of polynomials of N_i and the space of operators above. But note that the sum above can also accommodate operators where a particular frequency, say ω_i, does not appear simply by setting $p_i = q_i = 0$.

Now in a typical equilibrium state, we see that the only nonzero expectation values are products of N_i. This implies that

$$\langle \tilde{a}^\dagger \tilde{a}\rangle = \sum |A(p_1, q_1 \ldots p_n, q_n)|^2 \langle a_1^{q_1} (a_1^\dagger)^{p_1} a_1^{p_1} \times (a_1^\dagger)^{q_1} \ldots a_N^{q_N} (a_N^\dagger)^{p_N} a_N^{p_N} (a_N^\dagger)^{q_N}\rangle,$$

where the … indicate similar terms for all the other frequencies and cross terms vanish.

Evaluating the expectation value above in a state $|N_1 \ldots N_N\rangle$ we find that

$$\langle \tilde{a}^\dagger \tilde{a}\rangle = \sum_{p_j, q_j} |A(p_1, q_1 \ldots p_n, q_n)|^2 (N_1+1)_{q_1} \times (N_1 + q_1 - p_1 + 1)_{p_1} \ldots (N_N+1)_{q_N} \times (N_N + q_N - p_N + 1)_{p_N},$$

where the Pochhammer symbol is $(x)_n \equiv x(x+1)\ldots(x+n-1)$.

Next we notice that

$$[\tilde{a}, a_j] = -\sum A(p_1, q_1, \ldots p_n, q_n) q_j a_1^{p_1} \times (a_1^\dagger)^{q_1} \ldots a_j^{p_j} (a_j^\dagger)^{q_j - 1} \ldots a_N^{p_N} (a_N^\dagger)^{q_N}.$$

$$[\tilde{a}, a_j^\dagger] = \sum A(p_1, q_1, \ldots p_n, q_n) p_j a_1^{p_1} (a_1^\dagger)^{q_1} \ldots a_j^{(p_j - 1)} \times (a_j^\dagger)^{q_j} a_N^{p_N} \ldots (a_N^\dagger)^{q_N}.$$

Defining a new function, by the recursion relations

$$B(p_1, q_1 \ldots p_j, q_j, \ldots p_n, q_n) = (p_j + 1) A(p_1, q_1, \ldots p_j + 1, q_j, \ldots p_n, q_n),$$

$$B(p_1, q_1 \ldots p_j, q_j, \ldots p_n, q_n) = (q_j + 1) A(p_1, q_1, \ldots p_j, q_j + 1, \ldots p_n, q_n),$$

we see that we have

$$\sum_j \langle |[\tilde{a}, a_j]|^2\rangle + \langle |[\tilde{a}, a_j^\dagger]|^2\rangle = \sum [|B(p_1, q_1 \ldots p_n, q_n)|^2 (N_1 + q_1 - p_1 + 1)_{p_1} \times (N_1 + 1)_{q_1} \ldots (N_N + q_N - p_N + 1)_{p_N} (N_N + 1)_{q_N}].$$

In this case, we do not have a simple result like that of the simple harmonic oscillator. Indeed for some operators $\tilde{a}$ that are comprised of creation and annihilation operators, which have a very high occupancy in the state, it seems possible to make $\langle \tilde{a}^\dagger \tilde{a}\rangle \gg \langle \sum_j \langle |[\tilde{a}, a_j]|^2\rangle + \langle |[\tilde{a}, a_j^\dagger]|^2\rangle$. However, in most configurations and for almost all operators $\tilde{a}$, these two terms are comparable.

Note that in order to build an entire effectively isomorphic commuting algebra, we need a $\tilde{a}$ operator for each ordinary operator. Therefore even if, in some states, some of these operators have a small commutator with the ordinary operators, it is clear that there is not enough space in this chain of simple harmonic oscillators to accommodate mirror operators for each oscillator.

It is this intuition that carries over to the CFT. Consider the set of modes of generalized free fields. For simplicity, imagine separating them in frequency by ω_0, so that these modes all appear to be $\mathcal{O}_{n\omega_0, m}$. As usual, there could be other GFFs, while we are displaying only one of them. The main observation is the following. By putting a cutoff at the stretched horizon, we can limit the maximum angular momentum m that can appear for a given $\omega_n = n\omega_0$. Second, as we take $\omega_0 \propto \frac{1}{N^\alpha}$, where the precise power α depends on how we impose the cutoff above, then we find that these modes are already enough to account for the entropy of the CFT. (This is similar to the "brick wall" explanation of the black hole entropy in flat space [45].) Dimension counting, and the intuition from the simple harmonic oscillator above, would then suggest that there are no operators $\tilde{\mathcal{O}}_{\omega_n, m}$ that commute with all these modes.

While this commutator argument is a powerful constraint in practice, and was an important guiding principle in our construction [8,9], as the reader will notice it is hard to make it rigorous beyond this level. Moreover, power law suppressed commutators may be justified and even needed on physical grounds since the fields in the bulk are not strictly local. If we are willing to accept these small commutators, then the "commutator argument" above loses its power somewhat. For example, the reader can consult the talk [46] for an example that predates [8,9] and explores a model with such commutators.

This concludes our summary of the arguments that suggest that $\tilde{\mathcal{O}}_{\omega_n, m}$ cannot be found as state-independent operators in the CFT. A logical possibility is to accept that black holes have no interior. However, we believe that a more compelling alternative is that the black hole interior is described by state-dependent operators in the CFT.

VI. PARADOXES FOR THE ETERNAL BLACK HOLE

In this section, we show how versions of the paradoxes discussed in Sec. V also appear in the thermofield double state. It is sometimes believed, even by those who advocate that the single-sided black hole does not have an interior, that the thermofield double state nevertheless does correspond to an eternal black hole with a smooth horizon. For example, see [4].

We now show that this position is inconsistent. If we assume that the thermofield double state is dual to the eternal black hole, and demand only that the bulk theory respects diffeomorphism invariance—which is a minimal requirement in a theory of quantum gravity—then we can set up a large new class of states, all of which are dual to smooth black holes. This new class of states is obtained by performing one-sided diffeomorphisms on the geometry. We argue that diffeomorphisms that die off at the right boundary (but not, possibly, on the left boundary) should not affect the value of observables defined relationally from the right. This is a robust statement, and relies only on the fact that the gravity dual is diffeomorphism invariant—and not, in any way, on the equations of motion.

We then show that demanding that we find operators that behave correctly in *all* the states above leads to the same paradoxes that one finds in the single-sided case. Therefore a map between the bulk and the boundary, which can successfully describe the black hole interior in all these states, must be state dependent.

Our analysis is also useful because it indicates what state-dependence really means. To obtain the paradoxes above, we have to perform "extremely large" diffeomorphisms on one side—shifting the left boundary by time scales of order $e^{\mathcal{N}} \times \ell_{\rm AdS}$ before gluing it back to the geometry. What the analysis below shows is that it is not possible to use the same operator in the original state, and in all states that are obtained by deforming it with diffeomorphisms that could be exponentially large.

We start by reviewing the thermofield double state, and the geometry of the eternal black hole. Then we examine a class of "phase shifted" states, which are natural to consider from the point of view of the CFT, and show that they are also smooth because they are related to the original geometry by diffeomorphisms. We then set up analogues of the single-sided paradoxes. We defer the construction of *state-dependent* operators to Sec. VII.

A shorter version of the arguments of this section was also presented in [23]. In this section we elaborate on the arguments there and fill some gaps.

The paper [47] also discussed some subtleties of the map between the thermofield doubled state and the eternal black hole, and argued that the thermofield-eternal black hole duality is either incomplete or incorrect. While the argument presented here is similar to that of [47], our conclusion is different, as we show that if we consider a given state from (1) and small fluctuations about this state, then we can explicitly write down boundary operators that are dual to local bulk operators.

A. Review of the eternal black hole and the thermofield double

We start by reviewing the eternal black hole geometry and the duality proposed in [40]. The important point that we want to emphasize is the time reversal that is involved in gluing the geometry to the CFT, which is sometimes underemphasized.

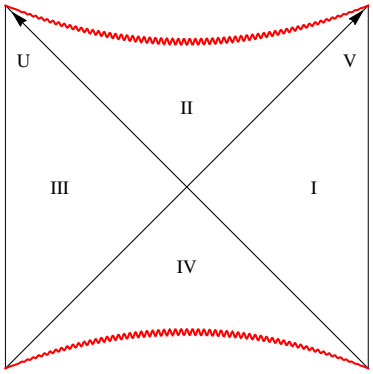


FIG. 7. Eternal black hole in AdS.

A schematic figure of the eternal black hole is shown in Fig. 7. For the eternal black hole, the metric is again given by (4.1) outside the horizon. Just as in Sec. IV A we introduce tortoise coordinates with the property that $r_* \to -\infty$ at the future horizon. The difference with the discussion in Sec. IV A is that after introducing the Kruskal coordinates, and extending the geometry inside the black hole we now extend the metric in a maximal way while assuming that there is no matter anywhere. This leads to the eternal black hole shown in Fig. 7, which also contains regions III–IV as shown in the figure. We can introduce Schwarzschild coordinates in all regions, and the relationship between the Kruskal and Schwarzschild coordinates is given below.

Region	Signs of (U,V)	Relationship to (t,r_*)
I	$U<0, V>0$	$U=-e^{\frac{2\pi}{\beta}(r_*-t)}, V=e^{\frac{2\pi}{\beta}(r_*+t)}$
II	$U>0, V>0$	$U=e^{\frac{2\pi}{\beta}(r_*-t)}, V=e^{\frac{2\pi}{\beta}(r_*+t)}$
III	$U>0, V<0$	$U=e^{\frac{2\pi}{\beta}(r_*-t)}, V=-e^{\frac{2\pi}{\beta}(r_*+t)}$
IV	$U<0, V<0$	$U=-e^{\frac{2\pi}{\beta}(r_*-t)}, V=-e^{\frac{2\pi}{\beta}(r_*+t)}$

(6.1)

The boundary, in these coordinates, is determined by the hyperbola $UV=-1$. On the other hand, the singularity lives at another hyperbola $UV=$ positive constant. The two null rays $U=0$, $V=0$ determine all four horizons. The horizon between region I and II, which would be the future horizon for the right-infalling observer, is at $U=0$. This same null ray also demarcates the boundary between regions IV and III and is therefore the "past horizon" for the left observer. The ray $V=0$ is the future horizon for the left-infalling observer, and the past horizon for the right observer.

The advantage of the choice of coordinates in (6.1) is that, in the UV plane, surfaces of $t=$ const are simply straight lines running through the origin. This includes the

horizons, which are $t=\infty$ and $t=-\infty$ respectively. Therefore, in these coordinates, geometrically we can think of time translations as "rotations" of the Kruskal diagram about the bifurcation point. Of course, we caution the reader that no finite rotation can rotate a line past the horizons. On the other hand, surfaces of constant r_* are hyperboloids that always stay within a single region.

Now, we mention an important point. When we associate the Schwarzschild time with the CFT time, we must "glue" the geometry to the left CFT with a flip in the time coordinate in region III. Therefore, denoting the time in CFT_R by t_R and the time in CFT_L by t_L we have the identifications

$$t_L=-t, \qquad t_R=t, \tag{6.2}$$

where t is the Schwarzschild time. An alert reader might ask, given that there is no natural choice of the origin of time, why one should not glue the geometry on the left as $t_L=-t+T$, where T is some constant. This is indeed possible, and is a central point in our discussion below.

We now turn to a description of the thermofield double state of the CFT. Maldacena conjectured [40] that the geometry we have described above is dual to an entangled state of *two* identical, noninteracting CFTs,

$$|\Psi_{\mathrm{tfd}}\rangle=\frac{1}{\sqrt{Z(\beta)}}\sum_E e^{-\frac{\beta E}{2}}\mathcal{T}|E,E\rangle. \tag{6.3}$$

Here $Z(\beta)$ is the partition function of a *single* CFT at the inverse temperature β and $|E,E\rangle\equiv|E\rangle_L\otimes|E\rangle_R$ is a tensor-product state of two energy eigenstates. Although the CFTs are entangled, they are noninteracting, and $\mathcal{T}$ is the time-reversal operator, which acts on left energy eigenstates.[16] The formula (6.3) is usually written with a tacit choice of the time-reversal operator

$$\mathcal{T}|E\rangle=|E\rangle,$$

in which case (6.3) reduces to the standard form

$$|\Psi_{\mathrm{tfd}}\rangle=\frac{1}{\sqrt{Z(\beta)}}\sum_E e^{-\frac{\beta E}{2}}|E,E\rangle.$$

We denote the Hamiltonian of the left CFT by $\boldsymbol{H}_L$ while we denote that of the right CFT by $\boldsymbol{H}$.[17]

We immediately see that $|\Psi_{\mathrm{tfd}}\rangle$ has a symmetry,

$$(\boldsymbol{H}_L-\boldsymbol{H})|\Psi_{\mathrm{tfd}}\rangle=0\Rightarrow e^{i(\boldsymbol{H}_L-\boldsymbol{H})T}|\Psi_{\mathrm{tfd}}\rangle=|\Psi_{\mathrm{tfd}}\rangle. \tag{6.4}$$

This symmetry of the thermofield double state corresponds to the isometry of the bulk geometry under $t\to t+T$. However, as is clear from the equation above, this symmetry corresponds to a shift in the CFT time in opposite directions in the two CFTs.

$$t\to t+T\Rightarrow t_R\to t_R+T; \qquad t_L\to t_L-T.$$

Now, let us examine why the eternal black hole, glued to the boundary as described above, is dual to the thermofield state $|\Psi_{\mathrm{tfd}}\rangle$, which involves a time-reversal on the left rather than a time-reversal combined with a time translation. Consider *mixed* correlators of a single trace operator in the thermofield state with one point (t_1,r_1,Ω_1) in region III and the other point (t_2,r_2,Ω_2) in region I. We would like to ensure that the bulk two-point function in this geometry has a limit that leads to these correlators.

$$\begin{aligned}&Z^2\lim_{r_1,r_2\to\infty}(r_1)^{\Delta}(r_2)^{\Delta}\langle\phi(t_1,r_1,\Omega_1)\phi(t_2,r_2,\Omega_2)\rangle_{\mathrm{EBH}}\\&\quad=\langle\Psi_{\mathrm{tfd}}|\mathcal{O}_1(-t_1,\Omega_1)\mathcal{O}_R(t_2,\Omega_2)|\Psi_{\mathrm{tfd}}\rangle,\end{aligned} \tag{6.5}$$

where the left-hand side is computed using bulk effective field theory in a metric that behaves asymptotically on both the right and the left as (4.1), and the right-hand side is computed as an expectation value in the thermofield state.

Computing the bulk two-point function in the eternal black hole metric is nontrivial, but we can do it patchwise as follows. We write down expansions for the field in regions I–III of the eternal black hole geometry. Only the near-horizon expansions are relevant and, with a short extension of the analysis of Sec. IV these expansions can be written as follows:

$$\begin{aligned}\phi(t,r_*,\Omega)&\overset{\substack{V>0\\U\to0^-}}{\to}\sum_m\int_0^\infty\frac{d\omega}{\sqrt{\omega}}a_{\omega,m}e^{-i\omega t}Y_m(\Omega)\\&\quad\times(e^{i\delta}e^{i\omega r_*}+e^{-i\delta}e^{-i\omega r_*})+\mathrm{H.c.}\end{aligned} \tag{6.6}$$

$$\begin{aligned}\phi(t,r_*,\Omega)&\overset{\substack{V>0\\U\to0^+}}{\to}\sum_m\int_0^\infty\frac{d\omega e^{-i\delta}}{\sqrt{\omega}}\\&\quad\times(a_{\omega,m}e^{-i\omega(t+r_*)}Y_m(\Omega)+\tilde{a}_{\omega,m}e^{i\omega(t-r_*)}Y_m^*(\Omega))+\mathrm{H.c.}\end{aligned} \tag{6.7}$$

$$\begin{aligned}\phi(t,r_*,\Omega)&\overset{\substack{U>0\\V\to0^+}}{\to}\sum_m\int_0^\infty\frac{d\omega e^{-i\delta}}{\sqrt{\omega}}\\&\quad\times(\tilde{a}_{L,\omega,m}e^{-i\omega(t+r_*)}Y_m(\Omega)+a_{L,\omega,m}e^{i\omega(t-r_*)}Y_m^*(\Omega))\\&\quad+\mathrm{H.c.}\end{aligned} \tag{6.8}$$

$$\begin{aligned}\phi(t,r_*,\Omega)&\overset{\substack{U>0\\V\to0^-}}{\to}\sum_m\int_0^\infty\frac{d\omega}{\sqrt{\omega}}a_{L,\omega,m}e^{i\omega t}Y_m(\Omega)\\&\quad\times(e^{i\delta}e^{i\omega r_*}+e^{-i\delta}e^{-i\omega r_*})+\mathrm{H.c.}\end{aligned} \tag{6.9}$$

[16]For simplicity, we assume that the CFT under consideration is invariant under time reversal and direct the reader to [48] for comments about the more general case.

[17]We use the notation $(\boldsymbol{H}_L,\boldsymbol{H})$ instead of what would be the more symmetric $(\boldsymbol{H}_L,\boldsymbol{H}_R)$ in order to keep the notation consistent with Sec. IX and also because we try to define right-relational observables, thus breaking the symmetry between the two CFTs.

Here we have introduced two new operators $a_{L,\omega m}$ and its mirror $\tilde{a}_{L,\omega,m}$. At the horizon between region III and region II, the field is defined using a left-relational coordinate system using the techniques of (3.1.1) and at the horizon between region I and II, it is defined using a right-relational coordinate system as usual.

The phase factors of $e^{i\delta}$ in the expansion above are slightly subtle. In (6.6) the two phase factors are fixed by the behavior of the mode at infinity by demanding (6.5) and by scattering in the bulk. In (6.7) the factor of $e^{-i\delta}$ multiplying the left mover is fixed but we have a choice of convention for the right movers. In region IV we have the same geometry but time reversed and this fixes the phase factors in (6.9) once again. We once again have some freedom in (6.8) for left-relational mirror.

Now notice that (6.7)–(6.8) have an overlapping regime of validity near the bifurcation point. Imposing the condition for the regularity of the two-point function that was discussed in Sec. IV we find that we must have

$$\langle a_{\omega,m} a_{L,\omega',m'}\rangle = \frac{e^{-\frac{\beta\omega}{2}}}{1-e^{-\beta\omega}}\delta(\omega-\omega')\delta_{mm'}.$$

Since the two-point function of the generalized free fields is the same in both CFTs, we can assume that (4.19) holds on both sides after we appropriately discretize the CFT modes. Therefore, from the bulk geometry and from (6.5) and after taking (6.2) into account we find that from the bulk we obtain the prediction for the boundary two-point function

$$\langle \Psi_{\rm tfd}|\mathcal{O}_{\omega_n,m}\mathcal{O}_{L\omega_n,m}|\Psi_{\rm tfd}\rangle = e^{-\frac{\beta\omega_n}{2}}G_\beta(\omega_n,m). \tag{6.10}$$

Note that here we have used a relationship between the boundary two-point function $G_\beta(\omega_n,m)$ and the boundary commutator $C_\beta(\omega_n,m)$ that appears in (4.19). This follows from the Kubo-Martin-Schwinger condition and is reviewed in [7].

To prove this we allow the matrix elements of these operators to be c_{ji} so that

$$\mathcal{O}_{\omega_n,m}\sum_i e^{-\frac{\beta E_i}{2}}|E_i,E_i\rangle = \sum_{i,j} e^{-\frac{\beta E_i}{2}}c_{ji}|E_i,E_j\rangle. \tag{6.11}$$

If the time-reversal symmetry acts as $\mathcal{T}|E\rangle = |E\rangle$ then using the fact that $\mathcal{T}\mathcal{O}_{\omega_n,m}\mathcal{T} = \mathcal{O}_{\omega_n,m}$, it follows that the c_{ji} must be real. Therefore

$$\mathcal{O}_{L\omega_n,m}\sum e^{-\frac{\beta E_j}{2}}|E_j,E_j\rangle = \sum e^{-\frac{\beta E_j}{2}}c_{ij}|E_i,E_j\rangle.$$

Since the matrix elements of c_{ji} are concentrated around $E_i - E_j = \omega_n$ we see that this is indeed true in the CFT because we can show that

$$\mathcal{O}_{L\omega_n,m}|\Psi_{\rm tfd}\rangle = e^{-\frac{\beta\omega_n}{2}}\mathcal{O}^\dagger_{\omega_n,m}|\Psi_{\rm tfd}\rangle.$$

From here (6.10) follows automatically.

We have therefore shown that the thermofield double state corresponds to the eternal black hole geometry glued with the specific identification (6.2). We return to this question below. We see that states with different correlators between the left and right boundary can also correspond to smooth geometries, albeit ones which are glued differently to the boundary.

B. Time-evolved thermofield states

We start by examining the effect of time evolution on the thermofield state. We consider the state

$$|\Psi_{\rm T}\rangle = e^{i(\boldsymbol{H}_L+\boldsymbol{H})\frac{T}{2}}|\Psi_{\rm tfd}\rangle = e^{i\boldsymbol{H}_L T}|\Psi_{\rm tfd}\rangle. \tag{6.12}$$

This is obtained by performing Hamiltonian evolution on the base thermofield state. We now perform both a geometric and a CFT analysis of these states. Our main results about these states come from understanding their geometry, as we do in the next subsection. However, we then provide some supporting arguments for these conclusions directly from the CFT.

1. Geometric analysis of time-shifted states

The action of the global symmetry group of the theory (which includes the Hamiltonian, of course) has been the subject of significant analysis in the general relativity literature [49]. The reader may find it useful to recall the analysis of Brown and Henneaux [50] who used such diffeomorphisms to analyze the action of the conformal group on the AdS_3 vacuum. For some more recent applications see [51]. The point is that Hamiltonian evolution—or evolution by some other global charge—corresponds to *large diffeomorphisms*. These operations may change the state of the theory.

A quick way to see this is as follows. Consider a nice slice that runs through the interior of the black hole and is anchored at the points (t_L,t_R). According to the standard analysis of the Hamiltonian constraint [25], the bulk Hamiltonian (including that of gravity and the other matter fields) must satisfy $H_{\rm bulk}|\Psi_{\rm tfd}\rangle = 0$. Therefore, time evolution of this slice is generated only by the boundary Hamiltonians $\boldsymbol{H}$ and $\boldsymbol{H}_L$. The action of $e^{i\boldsymbol{H}_L T}$ evolves this slice to another slice that is anchored at (t_L+T,t_R). This is shown in Fig. 8.

To summarize the geometric action of the left and right Hamiltonians is as follows.

(1) $e^{i\boldsymbol{H}_L T}\leftrightarrow$ large diffeomorphisms that die off at the right boundary, but not at the left boundary. On the left boundary, these diffeomorphisms shift points by $(t_L,\Omega_L)\to(t_L+T,\Omega_L)$.

(2) $e^{i\boldsymbol{H}T}\leftrightarrow$ large diffeomorphisms that die off at the left boundary, but not on the right boundary. On the right boundary, these diffeomorphisms shift points by $(t_R,\Omega_R)\to(t_R+T,\Omega_R)$.

We emphasize two important points. First, note that the operation $e^{i\boldsymbol{H}_L T}$ does not correspond to a unique

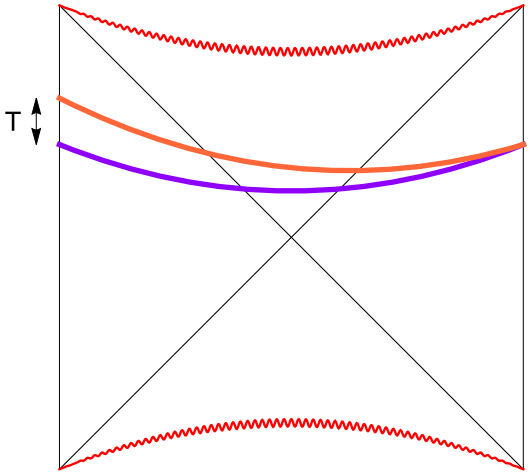


FIG. 8. The action of $e^{i\mathbf{H}_L T}$ is a large diffeomorphism that does not vanish on the left boundary. Its action on one nice slice is shown above.

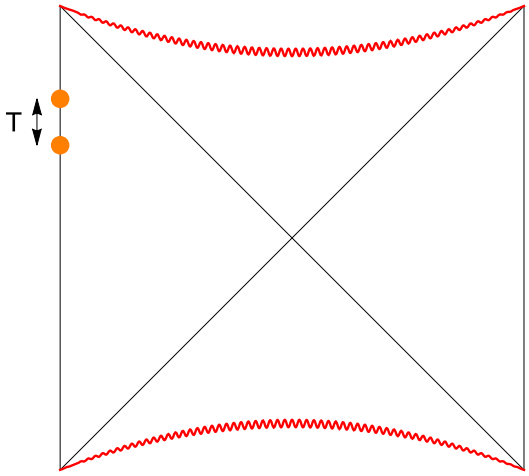


FIG. 9. Another diffeomorphism in the equivalence class of the diffeomorphism of Fig. 8: it slides points on the boundary but acts trivially in the bulk. This can be achieved by composing the diffeomorphism of Fig. 8 with a trivial diffeomorphism that cancels its action everywhere except for a region that is infinitesimally close to the boundary.

diffeomorphism. Rather there is an *equivalence class* of diffeomorphisms, all of which have the property outlined above. All diffeomorphisms in this equivalence class differ by *trivial diffeomorphisms*, which are those that die off at both boundaries. In terms of the nice slice picture of Fig. 8, this corresponds to the fact that we can choose to extend the nice slice in any way we like in the bulk, and a particular choice of nice slices is related to a choice of gauge. The left Hamiltonian must nevertheless evolve these slices forward in time. It achieves this because its Dirac brackets with operators in the interior depend on the choice of gauge. Therefore gauge-invariant statements about the diffeomorphism can only make reference to its action on the boundary and not in the interior.

Second, from the CFT we can see that while $e^{i\boldsymbol{H}_L T}$ and $e^{i\boldsymbol{H}T}$ change the state, an operation by $e^{i(\boldsymbol{H}_L-\boldsymbol{H})T}$ leaves the thermofield state invariance, since it satisfies $(\boldsymbol{H}_L-\boldsymbol{H})|\Psi_{\rm tfd}\rangle = 0$. Geometrically, this has the following meaning. Apart from the form of the metric itself, the thermofield state also has an additional piece of information that specifies the *relative placement* of the two boundaries. More specifically, there is an entire class of states—all of which correspond to the same gauge-invariant geometric quantities—which differ in how the left boundary is glued to the geometry.

To make this more precise, we describe a specific element of the class of diffeomorphisms that induces the action of $e^{i\boldsymbol{H}_L T}$. In the Kruskal coordinates U, V described above, we consider the following diffeomorphism $U \to U_T$, $V \to V_T$, where U_T, V_T are defined by

$$U_T = U(e^{\frac{2\pi T}{\beta}}\hat{\theta}(U-V) + \hat{\theta}(V-U)),$$
$$V_T = V(e^{-\frac{2\pi T}{\beta}}\hat{\theta}(U-V) + \hat{\theta}(V-U)),$$

where $\hat{\theta}(x)$ is an infinitely differentiable version of the theta function with the property that

$$\hat{\theta}(x) = \begin{cases} 1 & x > \epsilon \\ 0 & x < -\epsilon. \end{cases}$$

In the intermediate region $-\epsilon \le x \le \epsilon$ we can take f to be any smooth interpolating function between 0 and 1. For example, a function that satisfies all these criteria is given by

$$\hat{\theta}(x) = \frac{\theta(x+\epsilon)}{1+\theta(\epsilon-x)e^{\frac{\epsilon}{\epsilon+x}+\frac{\epsilon}{x-\epsilon}}}.$$

Since this is just a diffeomorphism, it does not actually change any gauge-invariant quantity that we can calculate in the bulk geometry. The correct way to picture the gauge-invariant effects of this diffeomorphism is to think of it as one that *slides* the left boundary by an amount T. Figure 9 may help the reader think of the effect of this diffeomorphism which, as we emphasized above, just changes the relation between the bulk and the boundary.

It is clear from the analysis above that the states $|\Psi_{\rm T}\rangle$ are also smooth states. This is an exact statement that does not rely on the bulk equations of motion and should be respected in any theory of quantum gravity that is diffeomorphism invariant. In particular, this implies that even for very large T, such as $T = e^{\mathcal{N}}$, the geometry remains smooth.

Time-shifted states for an infalling observer: Consider the experience of an infalling observer in the time-shifted thermofield state. This observer starts from region I, and falls towards the singularity. For example, such an observer could measure CFT correlators

$$\langle\Psi_{\rm T}|\boldsymbol{\phi}(t_1, r_1, \Omega_1)\ldots\boldsymbol{\phi}(t_n, r_n, \Omega_n)|\Psi_{\rm T}\rangle,$$

where all the points along his trajectory are defined relationally with respect to the right boundary as in Sec. III A 1.

We consider the relational observables, and the mirror creation and annihilation operators a little more carefully in the next subsection. However, for now we note an important property of the unshifted, standard thermofield state $|\Psi_{\rm tfd}\rangle$: if the observer jumps "earlier" or "later" in $|\Psi_{\rm tfd}\rangle$, according to the classical geometry, he will measure the same correlators. As the reader can verify, using classical geometry and quantum field theory quantized around this geometry we have

$$\langle\Psi_{\rm tfd}|\boldsymbol{\phi}(t_1,r_1,\Omega_1)\ldots\boldsymbol{\phi}(t_n,r_n,\Omega_n)|\Psi_{\rm tfd}\rangle = \langle\Psi_{\rm tfd}|\boldsymbol{\phi}(t_1+T,r_1,\Omega_1)\ldots\boldsymbol{\phi}(t_n+T,r_n,\Omega_n)|\Psi_{\rm tfd}\rangle.$$

Next, we note that

$$|\Psi_{\rm T}\rangle = e^{i\boldsymbol{H}_L T}|\Psi_{\rm tfd}\rangle = e^{i\boldsymbol{H}T}|\Psi_{\rm tfd}\rangle.$$

This results from the isometry (6.4) of the eternal black hole. So

$$\begin{aligned}\langle\Psi_{\rm tfd}|e^{-i\boldsymbol{H}_L T}\boldsymbol{\phi}(t_1,r_1,\Omega_1)\ldots\boldsymbol{\phi}(t_n,r_n,\Omega_n)e^{i\boldsymbol{H}_L T}|\Psi_{\rm tfd}\rangle &= \langle\Psi_{\rm tfd}|e^{-i\boldsymbol{H}T}\boldsymbol{\phi}(t_1,r_1,\Omega_1)\ldots\boldsymbol{\phi}(t_n,r_n,\Omega_n)e^{i\boldsymbol{H}T}|\Psi_{\rm tfd}\rangle\\ &= \langle\Psi_{\rm tfd}|\boldsymbol{\phi}(t_1-T,r_1,\Omega_1)\ldots\boldsymbol{\phi}(t_n-T,r_n,\Omega_n)|\Psi_{\rm tfd}\rangle.\end{aligned}$$

Therefore, if we combine the isometry of the eternal black hole with the fact that an infalling observer from the right observes the same geometry whenever he jumps in, then we obtain the same conclusion: the states $|\Psi_{\rm T}\rangle$ are smooth for all times. This is a second method to reach the conclusion that we already reached above. We now discuss these states from the perspective of the CFT.

2. CFT analysis of time-shifted states

We emphasize that the statement that we have made above—namely that the eternal black hole geometry should appear to be smooth under arbitrarily large diffeomorphisms—could be considered to be rather strong. Since we do not usually make statements about quantities that are exponentially large, using the geometry, let us understand these time-shifted states directly from the CFT.

The point we are making above is equivalent to the assertion that there is *no natural common origin of time* for the two CFTs. Usually, the origin of time is not relevant to any experiment. On the right CFT, for example, we declare some point in time to be $t=0$, pick some basis of operators that we can measure at that time, which we denote by $\mathcal{O}(0,\Omega)$ and declare that these are the Schrödinger operators. We can then classify states, using the eigenstates of these operators.

In our case, we have two CFTs. Roughly speaking, the original thermofield state involves entanglement between $\mathcal{O}(0,\Omega)$ and $\mathcal{O}_L(0,\Omega)$. The relation

$$\langle\Psi_{\rm tfd}|\mathcal{O}(0,\Omega)\mathcal{O}_L(0,\Omega')|\Psi_{\rm tfd}\rangle = \langle\Psi_{\rm T}|\mathcal{O}(0,\Omega)\mathcal{O}_L(T,\Omega')|\Psi_{\rm T}\rangle$$

tells us that the shifted states involve entanglement between $\mathcal{O}(0,\Omega)$ and $\mathcal{O}_L(T,\Omega)$. We can make an even stronger statement, as follows. Let us consider eigenstates of the Schrödinger picture operators which satisfy

$$\begin{aligned}\mathcal{O}(0,\Omega)|O_L(\Omega),O(\Omega)\rangle &= O(\Omega)|O_L(\Omega),O(\Omega)\rangle,\\ \mathcal{O}_L(0,\Omega)|O_L(\Omega),O(\Omega)\rangle &= O_L(\Omega)|O_L(\Omega),O(\Omega)\rangle,\end{aligned}$$

where $O_L(\Omega)$, $O(\Omega)$ are c-number functions that specify the eigenstate. We have a corresponding basis of eigenstates for the time-shifted Schrödinger basis operators, which are given by

$$\begin{aligned}\mathcal{O}(0,\Omega)|O_L(\Omega),O(\Omega)\rangle_T &= O(\Omega)|O_L(\Omega),O(\Omega)\rangle_T,\\ \mathcal{O}_L(T,\Omega)|O_L(\Omega),O(\Omega)\rangle_T &= O_L(\Omega)|O_L(\Omega),O(\Omega)\rangle_T.\end{aligned}$$

Then the thermofield state and the time-shifted thermofield state are identical when considered as wave functions on these states,

$$\langle\Psi_{\rm tfd}|O_L(\Omega),O(\Omega)\rangle = \langle\Psi_{\rm T}|O_L(\Omega),O(\Omega)\rangle_T.$$

So, unless we have some means of preferentially choosing the states $|O_L(\Omega),O(\Omega)\rangle$ over the states $|O_L(\Omega),O(\Omega)\rangle_T$, we must treat both the thermofield state and the time-shifted thermofield state on the same footing.

One distinguishing principle that is sometimes invoked in problems of this kind is to appeal to the "environment." We could state that the environment picks out the operators $\mathcal{O}_L(0,\Omega)$ and distinguishes them from the operators $\mathcal{O}_L(T,\Omega)$. However, this would tacitly break the time-translational invariance on the boundary. Moreover, from the point of view of gravity this would be very unusual; we would like the two coupled CFTs to autonomously describe the bulk geometry, and it would be unusual if some tacit reference to an external environment was important for deciding whether the geometry was smooth or not.

Let us consider some other methods that appear to uniquely pick the thermofield state but, on closer inspection, do not actually do so.

Euclidean path integral: The thermofield state can be defined by a Euclidean path integral on an interval of length β. More precisely we specify

$$\langle\Psi_{\rm tfd}|O_L(\Omega),O(\Omega)\rangle = \int_{O(0,\Omega)=O_L(\Omega)}^{O(\beta,\Omega)=O(\Omega)} e^{-S}[\mathcal{D}O],$$

where we have used $[\mathcal{D}O]$ to schematically represent the measure over fields in the theory, and placed boundary conditions so that, at time 0, the field is in the state specified by $O_L(\Omega)$ and at Euclidean time β it is in the state $O(\Omega)$. However, we see immediately that while the path integral on the right side has an unambiguous value, the interpretation of the path integral as a wave function on the left requires us to choose an origin of time. We could also write

$$\langle \Psi_{\rm T}|O_L(\Omega), O(\Omega)\rangle_T = \int_{O(0,\Omega)=O_L(\Omega)}^{O(\beta,\Omega)=O(\Omega)} e^{-S}[\mathcal{D}O].$$

So, using the Euclidean path integral to define the wave function begs the question of whether we should privilege $|O_L(\Omega), O(\Omega)\rangle_T$ versus the states $|O_L(\Omega), O(\Omega)\rangle$.

Time-reversal invariance: Another ostensible method of choosing the phases is to use invariance under the time-reversal operation. If we define the time-reversal operator in the left CFT as $\mathcal{T}|E\rangle = |E\rangle$, then the thermofield state is the only one of the family of time-shifted states that satisfies

$$\mathcal{T}|\Psi_{\rm tfd}\rangle = |\Psi_{\rm tfd}\rangle.$$

For the other states, recalling that the time-reversal operator acts antilinearly, we have

$$\mathcal{T}|\Psi_{\rm T}\rangle = |\Psi_{\rm -T}\rangle.$$

However, it is clear that this time-reversal operator itself involves the choice of an origin of time. We could just as well define a new time-reversal operation by a shift of the time-reversal above and a time translation. On the basis of energy eigenstates, we define

$$\mathcal{T}^T|E\rangle = e^{2iET}|E\rangle,$$

and extend this operation antilinearly on linear combinations of energy eigenstates. It is clear that

$$\mathcal{T}^T|\Psi_{\rm T}\rangle = |\Psi_{\rm T}\rangle.$$

The new operator $\mathcal{T}^T$ is as valid a time-reversal operator as the operator $\mathcal{T}$. Therefore, the idea that time-reversal invariance picks a particular origin of time is also specious; it can only do so if the origin of time is built into the time-reversal operator.

Time-shifted states as phase-modified states: We now turn to another property of the time-shifted states. This property is again suggestive of the fact that nothing very special happens if we take a long time limit of the time translation. Note that we can write the time-shifted states as

$$|\Psi_{\rm T}\rangle = e^{i\boldsymbol{H}_L T}|\Psi_{\rm tfd}\rangle = \frac{1}{\sqrt{Z(\beta)}}\sum_E e^{-\frac{\beta E}{2}} e^{i\phi_E}|E,E\rangle, \quad (6.13)$$

where ϕ_E are real phases. Since we expect the spectrum of the CFT to be chaotic at the high energies that dominate the state (6.13), we can obtain *almost any* choice of phases ϕ_E by choosing a suitable time translation. The relevant equation that we need to satisfy is

$$ET \bmod 2\pi = \phi_E,$$

and we can satisfy this to arbitrary accuracy for a chaotic collection of energies, if we are allowed to choose T from a large enough range.

There are some exceptions to the kinds of phases we can generate. For example, the energies of supersymmetric states are quantized integrally, and therefore we cannot choose their phases all independently. However, the set of supersymmetric states constitutes an *exponentially unimportant* subset in the thermofield state $|\Psi_{\rm tfd}\rangle$. More importantly, the energies within a conformal representation are integrally quantized. Therefore, by time evolution with the Hamiltonian,[18] we can only generate phases that satisfy

$$\phi[E] - \phi[E+1] = \phi[E+1] - \phi[E+2] \mod 2\pi.$$

The statement that there is no natural common origin of time translates, in this language, to the statement that there is no natural choice of phases for the energy eigenstates on both sides. (This is subject, of course, to the relations above.) The advantage of thinking in this language is that it is clear that the phases do not have any special behavior at late times. Therefore if we accept the standard interpretation that $e^{i\boldsymbol{H}_L T}$ acts as a large diffeomorphism in the bulk, for O(1) times, and preserves a smooth geometry, then it is natural to expect that this also happens for arbitrarily long T.

We caution the reader however that the argument above is a "naturalness" argument. It is predicated on the assumption that a natural bulk to boundary map should not privilege one pattern of random phases [obtained by translations of O(1)] from another pattern of random phases [obtained by translations of $O(e^{\mathcal{N}})$]. So it is suggestive and not a proof.

C. Relational observables in time-shifted states

We now turn to a detailed discussion of relational observables in time-shifted states. These operators are particularly important in our discussion of the eternal black hole.

We have already carefully defined relational observables in Sec. III A 1. The key point is as follows. These observables are defined relationally with respect to the right boundary. Therefore, if we consider diffeomorphisms that die off at the right boundary, then right-relational observables are invariant under such diffeomorphisms,

[18]The reader might notice that we can generate a slightly more general class of phases using other diffeomorphisms, such as those that rotate the S^{d-1}, but this is not relevant to our discussion.

even if the diffeomorphisms do not die off at the left boundary.

This point may be slightly confusing if one thinks of diffeomorphisms that shift the left boundary as acting everywhere in the spacetime. However, as we pointed out, these diffeomorphisms belong to an equivalence class, and a limiting element of the class is the diffeomorphism that simply "slides" the left boundary up and down while leaving the rest of the geometry invariant. If we consider this element of the class, it is clear that right-relational observables are left invariant.

Let us check this more explicitly by carefully repeating the derivation of Sec. III A 1. We start by defining points in the bulk as intersection points of null geodesics which end on the boundary. We introduce asymptotically AdS coordinates, so near the boundary the metric coincides with (3.12). These coordinates are (t, ρ, Ω) and the boundary is at $\rho = 1$. We now consider two solutions to the geodesic differential equation parametrized by ordinary AdS time (not necessarily an affine parameter) with the property that

$$\begin{aligned} &\vec{x}_1(t_1) = (t_1, \rho = 1, \Omega_1); \qquad \dot{\vec{x}}_1(0) = (1, -1, 0), \\ &\vec{x}_2(t_1 + \tau) = (t_1 + \tau, \rho = 1, \Omega_1); \quad \dot{\vec{x}}_2(t_1 + \tau) = (1, 1, 0). \end{aligned} \tag{6.14}$$

We then tune Ω_1 so that the geodesics meet. Given a particular value of t_1, $\Omega_1(t_1)$, we vary $\Omega_2(t_1 + \tau)$ so that the geodesics intersect at some t_i with $t_1 < t_i < t_1 + \tau$,

$$\rho_2(t_i) = \rho_1(t_i); \qquad \Omega_2(t_i) = \Omega_1(t_i),$$

and we denote the intersection point by $\vec{P}_i(t_1, \Omega_1, \tau)$ as in Sec. III A 1.

Let us now make a large diffeomorphism that dies off at the right boundary,

$$\vec{x} \to \vec{\xi}(\vec{x}). \tag{6.15}$$

To implement this diffeomorphism in a quantum field theory, we can act on all *fields* (including the metric), rather than points, with the inverse transformation. The new scalar fields $\bar{\phi}(\vec{x})$ are given by

$$\bar{\phi}(\vec{x}) = \phi(\vec{\xi}^{-1}(\vec{x})).$$

The action of the diffeomorphism on the metric is

$$g_{\bar{\mu}\bar{\nu}}(\vec{x}) \to \frac{\partial x^\mu}{\partial \xi^{\bar{\mu}}} \frac{\partial x^\nu}{\partial \xi^{\bar{\nu}}} g_{\mu\nu}(\vec{\xi}^{-1}(\vec{x})). \tag{6.16}$$

Now if we transform the entire geodesic trajectory specified by the solution to the geodesic equation with initial conditions (6.14) by means of the diffeomorphism (6.15), then we get a new trajectory that is a geodesic with respect to the new metric (6.16).

The boundary conditions (6.14) remain invariant under the diffeomorphism since, by assumption, ξ turns into the identity at the boundary. Moreover, if the original geodesics intersected, then the new geodesics also intersect. In particular the new intersection point, $\bar{P}_i$, is just given by the transform of the original intersection point

$$\vec{\bar{P}}_i(t_1, \Omega_1, \tau) = \vec{\xi}(\vec{P}_i(t_1, \Omega_1, \tau)),$$

where we are using the same notation as (3.16).

Now consider evaluating a scalar field at this intersection point. Clearly we have

$$\bar{\phi}(\vec{\bar{P}}_i) = \phi(\vec{\xi}^{-1}(\vec{\xi}(\vec{P}_i))) = \phi(\vec{P}_i),$$

which is the same value as it had before the diffeomorphism. Therefore, scalar observables defined at points which are related relationally to the right boundary are invariant under left diffeomorphisms.

This logic extends to points behind the horizon. Recall that these points were defined by solutions to the geodesic equation, where the affine parameter was normalized by using the points outside the horizon already defined above. Clearly, in the new metric the new geodesics are again given by $\vec{\xi}(\vec{x}(\lambda))$, and by the same logic scalar variables evaluated inside the horizon are invariant under any diffeomorphism that dies off at the right boundary.

1. Commutator of mirror operators

Note that, in the analysis above, it was important that the boundary conditions (6.14) were not altered by the diffeomorphisms. If we consider diffeomorphisms that do not die off at the right boundary, then the right-relational observables do transform, but in a simple manner. Under a diffeomorphism that shifts points on the right boundary by $t_R \to t_R + T$, we have

$$\vec{\bar{P}}_i(t, \Omega, \tau) = \vec{\xi}(\vec{P}_i(t - T, \Omega, \tau)).$$

For the field operators, defined relationally with respect to the right boundary, this leads to

$$\begin{aligned} e^{iH_L T} \phi(t_R, \Omega, \lambda) e^{-iH_L T} &= \phi(t_R, \Omega, \lambda), \\ e^{iHT} \phi(t_R, \Omega, \lambda) e^{-iHT} &= \phi(t_R + T, \Omega, \lambda), \end{aligned} \tag{6.17}$$

where H_L and H are the left and right boundary Hamiltonians respectively.

We now write down a mode expansion for the fields in front of and behind the horizon, as in (6.6)–(6.7). The conditions (6.17) imply that when we try and find CFT operators that can play the role of these mirrors then they must have the CFT commutation relations

$$[\boldsymbol{H}, \boldsymbol{a}_{\omega,m}] = -\omega \boldsymbol{a}_{\omega,m}, \qquad [\boldsymbol{H}_L, \boldsymbol{a}_{\omega,m}] = 0,$$
$$[\boldsymbol{H}, \tilde{\boldsymbol{a}}_{\omega,m}] = \omega \tilde{\boldsymbol{a}}_{\omega,m}, \qquad [\boldsymbol{H}_L, \tilde{\boldsymbol{a}}_{\omega,m}] = 0. \tag{6.18}$$

We remind the reader that the asymmetry above arises because these are right-relational modes. The relation (6.18) must hold approximately within low point correlation functions, and not necessarily as operators. However, within correlators they are crucial to ensure that the field operators transform correctly under large diffeomorphisms.

We proceed to now argue that it is impossible to find state-independent operators $\tilde{a}_{\omega,m}$ that have the right properties to play the role of mirror operators behind the horizon in the entire family of time-shifted states.

D. Naive construction of local operators in the thermofield double

We start by considering the naive construction of local operators in the thermofield double. We show that this does not satisfy the conditions above and, therefore, cannot be correct. In particular we identify CFT operators $\tilde{\boldsymbol{a}}_{\omega_n,m}$ with the properties that we derived from the bulk above.

The naive construction of local operators proceeds by simply identifying discretized mirror modes with modes on the left CFT,

$$\tilde{\boldsymbol{a}}_{\omega_n,m} \underset{\text{naive}}{\longrightarrow} \boldsymbol{a}_{L\omega_n,m}.$$

However, this is clearly wrong as a computation of the two-point function across the horizon shows. If we now compute this two-point correlator in the time-shifted state, we find that

$$\langle \Psi_{\rm T}|\boldsymbol{a}_{L\omega_n,m}\boldsymbol{a}_{\omega_n,m}|\Psi_{\rm T}\rangle = e^{i\omega_n T}\frac{e^{-\frac{\beta\omega_n}{2}}}{1-e^{-\beta\omega_n}},$$
$$\langle \Psi_{\rm T}|\boldsymbol{a}^\dagger_{L\omega_n,m}\boldsymbol{a}^\dagger_{\omega_n,m}|\Psi_{\rm T}\rangle = e^{-i\omega_n T}\frac{e^{-\frac{\beta\omega_n}{2}}}{1-e^{-\beta\omega_n}}.$$

Let us call the CFT operator obtained by using this naive mode $\boldsymbol{\phi}^{\rm n}$. Now, repeating the computation of the two-point function that we performed in Sec. IV, with point 1 outside the horizon and point 2 behind the horizon we find that

$$\lim_{V_1-V_2\to 0}\langle \Psi_{\rm T}|\partial_U\boldsymbol{\phi}^{\rm n}(U_1,V_1,\Omega_1)\partial_U\boldsymbol{\phi}^{\rm n}(U_2,V_2,\Omega_2)|\Psi_{\rm T}\rangle$$
$$= c\frac{\delta^{d-1}(\Omega_1-\Omega_2)}{(U_1-U_2e^{-\frac{2\pi T}{\beta}})^2},$$
$$\lim_{U_1-U_2\to 0}\langle \Psi_{\rm T}|\partial_V\boldsymbol{\phi}^{\rm n}(U_1,V_1,\Omega_1)\partial_V\boldsymbol{\phi}^{\rm n}(U_2,V_2,\Omega_2)|\Psi_{\rm T}\rangle$$
$$= c\frac{\delta^{d-1}(\Omega_1-\Omega_2)}{(V_1-V_2)^2}, \tag{6.19}$$

where c is a normalization constant. Clearly this is not the correct result. In particular, the first line of (6.19) does not have the right behavior when $U_1 \to U_2$. We obtain a similar pathology by considering the boundary between region II and region III.

This was only to be expected since the operators $\boldsymbol{a}_{L\omega}$ clearly do not obey the correct commutators with the Hamiltonian that we demanded above. Therefore, it is incorrect to identify $\tilde{\boldsymbol{a}}_{\omega_n,m}$ with $\boldsymbol{a}_{L,\omega_n,m}$ as has been done commonly in the literature.

E. Paradoxes for the eternal black hole

We now set out various paradoxes, similar to the ones outlined by [2–4], which show that the relational observable defined above *cannot* be realized by a linear operator. These paradoxes were already outlined concisely in [23], and we suggest that the reader consult that paper alongside this section. Our arguments here are more detailed variants of the arguments there.

Let us assume that some state-independent operators $\tilde{\boldsymbol{a}}_{\omega_n,m}$ exist with the properties that we derived earlier. If so we can multiply them with the appropriate modes and construct state-independent operators $\boldsymbol{\phi}(U,V,\Omega)$ in the thermofield double state and in a right-relational gauge. Then, consider

$$C(U_1,V_1,\Omega_1,\ldots U_n,V_n,\Omega_n)$$
$$= \langle \Psi_{\rm T}|\boldsymbol{\phi}(U_1,V_1,\Omega_1)\ldots\boldsymbol{\phi}(U_n,V_n,\Omega_n)|\Psi_{\rm T}\rangle.$$

From the arguments above we have

$$\frac{d}{dT}C(U_1,V_1,\Omega_1,\ldots U_n,V_n,\Omega_n) = 0.$$

Second, from the discussion in Sec. III, we expect this T-independent answer to correspond to the correlators computed by effective field theory in the eternal black hole. This expectation is indicated in (3.10). Now, for any operator $\boldsymbol{A}_\alpha$ we have

$$\langle \Psi_{\rm T}|\boldsymbol{A}_\alpha|\Psi_{\rm T}\rangle = \frac{1}{Z(\beta)}\Big[\sum_E e^{-\beta E}\langle E,E|\boldsymbol{A}_\alpha|E,E\rangle$$
$$+\sum_{E'\neq E} e^{\frac{-\beta(E+E')}{2}}e^{i(E-E')T}\langle E',E'|\boldsymbol{A}_\alpha|E,E\rangle\Big].$$

Even if we know that this expectation value is T independent, we must be careful not to immediately discard the second term above. This is because, if $\boldsymbol{A}_\alpha$ happens to be an operator with support on narrowly separated eigenstates $E-E'=\mathrm{O}(e^{-\frac{S}{2}})$, then the time variation of the second term will be negligible and so it may appear to be time independent for short times. However, if we demand

$$\langle \Psi_{\rm T}|\boldsymbol{A}_\alpha|\Psi_{\rm T}\rangle = \langle \Psi_{\rm tfd}|\boldsymbol{A}_\alpha|\Psi_{\rm tfd}\rangle,$$

even for exponentially long times, then the contribution to the expectation value can only come from diagonal terms.

In the case of the correlator under consideration this implies that

$$\frac{1}{Z(\beta)}\sum_E e^{-\beta E}\langle E,E|\boldsymbol{\phi}(U_1,V_1,\Omega_1)\ldots\boldsymbol{\phi}(U_n,V_n,\Omega_n)|E,E\rangle = C(U_1,V_1,\Omega_1,\ldots U_n,V_n,\Omega_n).$$

Using the standard arguments from the equivalence of the canonical and the microcanonical ensemble this means that for a typical eigenstate pair $|E,E\rangle$ at the energy relevant to the eternal black hole

$$\langle E,E|\boldsymbol{\phi}(U_1,V_1,\Omega_1)\ldots\boldsymbol{\phi}(U_n,V_n,\Omega_n)|E,E\rangle = C(U_1,V_1,\Omega_1,\ldots U_n,V_n,\Omega_n).$$

At an intuitive level this is already a strange conclusion because the energy-eigenstate pair that appears above has *no entanglement*. We have shown above that no state-independent operators $\boldsymbol{\phi}(U,V,\Omega)$ can reproduce the effective field theory correlators in arbitrary *single-sided* energy eigenstates. How can such operators correctly reproduce this answer in *two-sided* eigenstate pairs?

We can turn this into a sharp contradiction as follows. In the eigenstate pair $|E,E\rangle$ with no entanglement, we expect that there is no geometric wormhole. Therefore no excitation generated by the left observer can affect the correlators observed by the right-infalling observer. In particular, if the left observer decides to act with an arbitrary unitary, $\boldsymbol{U}_L$, we should have

$$\langle E,E|\boldsymbol{U}_L^\dagger\boldsymbol{\phi}(U_1,V_1,\Omega_1)\ldots\boldsymbol{\phi}(U_n,V_n,\Omega_n)\boldsymbol{U}_L|E,E\rangle = \langle E,E|\boldsymbol{\phi}(U_1,V_1,\Omega_1)\ldots\boldsymbol{\phi}(U_n,V_n,\Omega_n)|E,E\rangle. \quad (6.20)$$

We can use this freedom to map the left energy eigenstate to some fixed state, $\boldsymbol{U}_L|E,E\rangle = |F,E\rangle$, where F could even correspond to the left CFT vacuum. This means that the operators $\boldsymbol{\phi}(U,V,\Omega)$ must reproduce the correct correlators in all states $|F,E\rangle$ and must be independent of F. This can only be if they are ordinary operators in the right CFT. But we have already proved that there are no state-independent operators in the right CFT. Therefore our starting assumption—that such operators exist in the doubled CFT—must be wrong.

The reader may consult [23] for concrete versions of the $N_a \neq 0$ argument, and the negative occupancy argument phrased directly in the doubled CFT. Here, we conclude by briefly reemphasizing the importance of (6.20), which states that there is no wormhole in eigenstate pairs.

In Sec. VII we review the construction of state-dependent operators in a single CFT that can correctly reproduce effective field theory correlators about a black hole. This construction was first described in [8,9]. Let us denote such operators acting only in the original (right) CFT, and defined about an energy eigenstate $|E\rangle$ by $\boldsymbol{\phi}^{\{E\}}(U,V,\Omega)$. The superscript E indicates that they reproduce the expected effective field theory answers when evaluated in correlators about $|E\rangle$ and reasonable excitations of this state. Now, consider the following state-independent operator, which acts in the Hilbert space of two CFTs,

$$\boldsymbol{\Theta}(U,V,\Omega) = \sum_E P_{E_L}\otimes\boldsymbol{\phi}^{\{E\}}(U,V,\Omega),$$

where P_{E_L} is the projector onto the energy eigenstate on the left, $P_{E_L}\equiv|E_L\rangle\langle E_L|$, and the sum is over all energy eigenstates.

Now $\boldsymbol{\Theta}(U,V,\Omega)$ has some interesting properties. When evaluated in the thermofield double, we find

$$\langle\Psi_{\rm tfd}|\boldsymbol{\Theta}(U_1,V_1,\Omega_1)\ldots\boldsymbol{\Theta}(U_n,V_n,\Omega_n)|\Psi_{\rm tfd}\rangle = \frac{1}{Z(\beta)}\sum_E e^{-\beta E}\langle E|\boldsymbol{\phi}^{\{E\}}(U_1,V_1,\Omega_1)\ldots\boldsymbol{\phi}^{\{E\}} \times(U_n,V_n,\Omega_n)|E\rangle. \quad (6.21)$$

Note that the sum on the right is in a single CFT since the P_{E_L} term simply makes cross terms vanish and gives 1 for the diagonal terms.

Since $\boldsymbol{\phi}^{\{E\}}(U,V,\Omega)$ is only evaluated in the state $|E\rangle$ and its excitations, the expression above does yield the answer expected from effective field theory. Note that $\boldsymbol{\Theta}(U,V,\Omega)$ also produces the following correlators about eigenstate pairs:

$$\langle E,E|\boldsymbol{\Theta}(U_1,V_1,\Omega_1)\ldots\boldsymbol{\Theta}(U_n,V_n,\Omega_n)|E,E\rangle = \langle E|\boldsymbol{\phi}^{\{E\}}(U_1,V_1,\Omega_1)\ldots\boldsymbol{\phi}^{\{E\}}(U_n,V_n,\Omega_n)|E\rangle.$$

Using the equivalence between the canonical and microcanonical ensemble, these correlators are approximately the same as the thermofield correlators in (6.21). These correlators would suggest that the geometry in eigenstate pairs, as seen by the right-infalling observer, is almost the same in eigenstate pairs as in the thermofield. While this conclusion is correct, as we see below, the operator $\boldsymbol{\Theta}(U,V,\Omega)$ cannot be the correct CFT operator dual to local bulk fields.

This is because $\boldsymbol{\Theta}(U,V,\Omega)$ violates the no wormhole condition and keeps the wormhole open even when there is no entanglement. In particular, using a left unitary that acts as $\boldsymbol{U}_L|E,E\rangle = |F,E\rangle$ we find that

$$\langle E,E|\boldsymbol{U}_L^\dagger\boldsymbol{\Theta}(U_1,V_1,\Omega_1)\ldots\boldsymbol{\Theta}(U_n,V_n,\Omega_n)\boldsymbol{U}_L|E,E\rangle = \langle E|\boldsymbol{\phi}^{\{F\}}(U_1,V_1,\Omega_1)\ldots\boldsymbol{\phi}^{\{F\}}(U_n,V_n,\Omega_n)|E\rangle.$$

But these are correlators of $\boldsymbol{\phi}^{\{F\}}(U,V,\Omega)$ evaluated about a different eigenstate and, in general, these lead to exponentially small answers. Therefore, $\boldsymbol{\Theta}(U,V,\Omega)$ cannot be the correct field operators in the eternal black hole because they would predict that even in eigenstate pairs, by performing the unitary transformation discussed above a left observer could alter the correlators of a right-infalling observer. So we see that condition (6.20) is important in ruling out such putative state-independent operators. In the next section, we show how the interior of the eternal black hole can be correctly constructed using state-dependent bulk to boundary maps.

Before concluding this section, we should mention that our arguments should be distinguished from those of [52,53], who suggested that the duality between the eternal black hole and the thermofield double does not hold. Although we do not engage with this in detail, we briefly indicate our point of disagreement. The authors of [52] suggested that there was an ambiguity in the duality between the thermofield double and the eternal black hole. In particular, they argued that the CFT cannot distinguish between this case and another bulk geometry where the bulk Hamiltonian has been modified by removing the interaction between the left and the right at the bifurcation point. Alternately, this corresponds to adding a delta-function source there in a manner that appears to be hidden from both CFTs. They argued that this leads to an ambiguity that invalidates the duality.

While this argument may have been plausible if the bulk theory had been an ordinary quantum field theory, it is inapplicable to a theory of quantum gravity. The Hamiltonian constraint rules out the alternate bulk Hamiltonian considered above. It is this crucial feature of the bulk that allows the boundary to know the details of the bulk Hamiltonian and allows the duality to be consistent.

VII. DEFINITION OF THE MIRROR OPERATORS

In the past sections, we have set up paradoxes that show that no state-independent operator can correctly satisfy the conditions outlined in Sec. IV. We have shown that these paradoxes apply to both the single-sided CFT and the thermofield double.

We now review and extend the definition of the mirror operators provided in [8,9]. These operators are state dependent. What this means, in our context, is as follows. Say that we are computing expectation values of a mirror operator within a correlation function

$$\langle\Psi|\mathcal{O}_{\omega_1,m_1}\ldots\tilde{\mathcal{O}}_{\omega_p,m_p}\ldots\mathcal{O}_{\omega_n,m_n}|\Psi\rangle,$$

where $|\Psi\rangle$ is an equilibrium state. Then the statement is that the operator $\tilde{\mathcal{O}}_{\omega,m}$ depends in a subtle manner on the sandwiching state $|\Psi\rangle$.

This would imply that when one speaks of local operators in gravity, or of their modes, then at least behind the horizon of a black hole it is important to specify the state that one is referring to. A given local operator is good to describe physics in a given state and in small excitations about that state. If we consider another microstate which is "far away," in the sense that it cannot be obtained from the original microstate by the action of a small number of single-trace operators, then we must use a different operator to describe the "same physical quantity."

In this section we first review the construction that we presented in [8,9] both for equilibrium and near-equilibrium states. We show how this completely resolves all the paradoxes of [2–4]. Our review will be brief, and we direct the reader to those papers for a more detailed exposition.

A significant new element in this paper is that we discuss the action of our operators on *superpositions* of states. This is important because we show that even though our operators are state dependent, the infalling observer does not observe any deviations from linearity for small superpositions of equilibrium or near-equilibrium states.

Next, we also describe the construction of mirror operators for the thermofield double and its time-shifted cousins. This construction can be obtained as a special case of our construction, as applied to an entangled state. However, in this section we also show how one could guess this solution independently. The analysis of (7.6) is useful because it helps to elucidate the nature of state-dependence.

A. The set of natural observables and the little Hilbert space about a state

Consider the modes of the generalized free field operators that were defined in (4.18). As we explained there, we have discretized these modes $\mathcal{O}_{\omega_n,m}$ both by selecting some discrete set of frequencies, and also by choosing a time band on the boundary that we integrate over to transform to frequency space.

We now consider the set of polynomials in these modes that we denote by

$$\mathcal{A}_{\rm gff} = \text{span}\{\mathcal{O}_{\omega_1,m_1}, \mathcal{O}_{\omega_1,m_1}\mathcal{O}_{\omega_2,m_2},\ldots, \mathcal{O}_{\omega_1,m_1}\mathcal{O}_{\omega_2,m_2}\ldots\mathcal{O}_{\omega_K,m_K}\}. \tag{7.1}$$

This means that this set comprises all monomials of the form displayed above, and arbitrary linear combinations of these monomials. In addition, we consider the set of polynomials—limited to small orders—in the CFT Hamiltonian.[19]

[19]For a more careful treatment of other conserved charges, including in cases where the CFT has a non-Abelian symmetry, we refer the reader to Sec. III B 4 of [9].

$$\mathcal{A}_{\boldsymbol{H}} = \text{span}\{\boldsymbol{H}, \boldsymbol{H}^2 \ldots \boldsymbol{H}^n\}.$$

We then consider the set of observables involving insertions of both the generalized free fields and the CFT Hamiltonian

$$\mathcal{A} = \mathcal{A}_{\text{gff}} \otimes \mathcal{A}_{\boldsymbol{H}}. \tag{7.2}$$

The dimension of this set is denoted by

$$\mathcal{D}_{\mathcal{A}} = \dim(\mathcal{A}).$$

We often refer to arbitrary elements of this set, comprising generalized free fields by

$$\boldsymbol{A}_\alpha \in \mathcal{A}_{\text{gff}}.$$

We emphasize by default that the notation $\boldsymbol{A}_\alpha$ *does not* include the CFT Hamiltonian. If we want to consider an element from $\mathcal{A}$ that might include $\boldsymbol{H}$, we state this explicitly.

We want to restrict $\mathcal{A}$ to be the set of reasonable experiments that one can perform in the bulk, and still expect to observe effective field theory about a given background. This excludes any monomial in (7.2) that has a very high total energy

$$\sum \omega_i \ll \mathrm{O}(\mathcal{N}).$$

Similarly, this also excludes any monomial that has a very large number of insertions. So

$$K \ll \mathrm{O}(\mathcal{N}),$$

for all monomials displayed in (7.2). These restrictions imply, as a consequence, that

$$\mathcal{D}_{\mathcal{A}} \ll \mathrm{O}(e^{\mathcal{N}}).$$

The set $\mathcal{A}$ is approximately an algebra because we can usually multiply two of its element to obtain another element. However, this is not always the case because of edge effects—where such a multiplication may take us beyond the cutoff we have imposed. In this paper we usually do not keep track of these edge effects.

The set of "reasonable operators" can be used to excite a state. This leads us to consider the space

$$\mathcal{H}_\Psi = \mathcal{A}|\Psi\rangle \equiv \text{span}\Big\{\sum \alpha_p \boldsymbol{A}_p |\Psi\rangle\Big\},$$

where $\boldsymbol{A}_\alpha$ may include $\boldsymbol{H}$. We denote the projector on this subspace by $\boldsymbol{P}_{\mathcal{H}_\Psi}$. The fact that $\mathcal{A}$ is approximately an algebra implies that we can consider the action of its elements as $\boldsymbol{A}_\alpha$: $\mathcal{H}_\Psi \to \mathcal{H}_\Psi$. This is subject to the same edge-effect caveat above.

We sometimes call the space $\mathcal{H}_\Psi$ the little Hilbert space about the space $|\Psi\rangle$, since it contains the part of the Hilbert space that is accessible within effective field theory. Conceptually, this little Hilbert space is very important. We show a schematic figure of this set in Fig. 10.

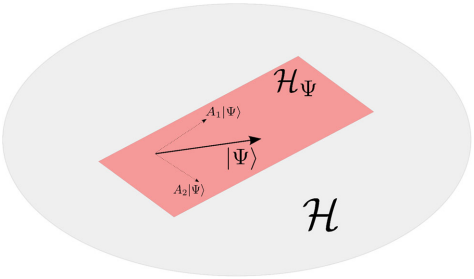


FIG. 10. A cartoon of the little Hilbert space $\mathcal{H}_\Psi$ as the relevant subspace in the full Hilbert space.

B. Equilibrium and near-equilibrium states

The next ingredient in our construction is the classification of states. First, we consider equilibrium states. Intuitively, these are states where a black hole in the bulk has not been disturbed for a long time. We then expect that all excitations both outside and inside the horizon have died off, leaving behind a smooth horizon and an empty interior. We now want to make this precise in the CFT.

Let us review some *necessary* conditions for us to classify a state as being in equilibrium. (As we discuss in Sec. VIII these conditions are not quite sufficient.) The first is that correlation functions in an equilibrium state should be invariant under time translation.

We consider the expectation value of an element of the set of observables $\boldsymbol{A}_p \in \mathcal{A}$, as a function of time. This is defined as

$$\chi_p(t) = \langle\Psi|e^{iHt}\boldsymbol{A}_p e^{-iHt}|\Psi\rangle, \tag{7.3}$$

where it is important that $\boldsymbol{A}_p$ may include $\boldsymbol{H}$. Intuitively, while there may be small fluctuations in this expectation value, we expect that in an equilibrium state, these fluctuations are extremely unlikely. The size of the fluctuations is measured by

$$\nu_p = \frac{1}{T_{\text{b}}}\int_0^{T_{\text{b}}} |(\chi_p(t) - \chi_p(0))|dt. \tag{7.4}$$

An estimate of these fluctuations [9] suggests that a state should be classified as being in equilibrium if

$$\nu_p = \mathrm{O}(e^{-\frac{S}{2}}), \quad \forall\ p. \tag{7.5}$$

Note that the definition requires this to hold for all observables in $\mathcal{A}$.

The condition for time independence of correlators can be imposed very accurately. However, this condition is necessary but not sufficient in order for us to apply our definition of the mirror operators. In particular, to apply our

definition, we would also like the state to correspond to a state at a single temperature. For example, consider the state $\frac{1}{\sqrt{2}}(|E_1\rangle + |E_2\rangle)$ where E_1, E_2 are two distinct energy eigenstates at substantially separated energies. For example, we could take $E_2 \approx 10E_1$. It is easy to verify, using the eigenstate thermalization hypothesis, that this state meets the criterion (7.5) above. However we think of this as a sum of two separate equilibrium states.

Now we describe near-equilibrium states. Near-equilibrium states are simply obtained by exciting an equilibrium state with an exponentiated Hermitian element of the set of observables $\mathcal{A}$.

$$|\Psi^{\rm ne}\rangle = \boldsymbol{U}|\Psi\rangle, \qquad \boldsymbol{U} = e^{i\boldsymbol{A}_p}, \qquad \boldsymbol{A}_p^\dagger = \boldsymbol{A}_p. \tag{7.6}$$

In [8,9], we showed that given a state $|\Psi^{\rm ne}\rangle$ of this kind, the decomposition into a unitary $\boldsymbol{U}$ and a base-equilibrium state $|\Psi\rangle$ was essentially unique. The reason for this is very simple. Given an equilibrium state $|\Psi\rangle$, if we excite it with a unitary we necessarily spoil the time-translational invariance criterion of (7.5). Therefore, given a state $|\Psi^{\rm ne}\rangle$, once we have found a decomposition (7.6) that works to make all correlators time-translationally invariant in the base state $|\Psi\rangle$, we know that it must be the right one.

C. Mirrors for equilibrium and near-equilibrium states

We now consider the definition of mirror operators for the states considered above. We start with an equilibrium state $|\Psi\rangle$ with inverse temperature β. First we consider excitations of this state with $\boldsymbol{A}_\alpha \in \mathcal{A}_{\rm gff}$. This set was defined in (7.1) and excludes the Hamiltonian. We now *define* mirror operators on this subspace of $\mathcal{H}_\Psi$ through the linear equations

$$\tilde{\boldsymbol{\mathcal{O}}}_{\omega_n,m}\boldsymbol{A}_\alpha|\Psi\rangle = e^{-\frac{\beta\omega_n}{2}}\boldsymbol{A}_\alpha\boldsymbol{\mathcal{O}}^\dagger_{\omega_n,m}|\Psi\rangle. \tag{7.7}$$

We can use this definition recursively to define the mirrors of products of operators as well,

$$\tilde{\boldsymbol{A}}_\alpha\boldsymbol{A}_\beta|\Psi\rangle = \boldsymbol{A}_\beta e^{-\frac{\beta H}{2}}\boldsymbol{A}_\alpha^\dagger e^{\frac{\beta H}{2}}|\Psi\rangle.$$

These relations specify the action of $\tilde{\boldsymbol{\mathcal{O}}}_{\omega_n,m}$ on $\mathcal{H}_\Psi$. The action of this operator outside this space is irrelevant for questions within effective field theory. We expect (7.7) to hold at leading order in $\frac{1}{\mathcal{N}}$.

However, we do specify its commutator with the Hamiltonian and this fixes some $\frac{1}{\mathcal{N}}$ corrections.

$$[\tilde{\boldsymbol{\mathcal{O}}}_{\omega_n,m}, \boldsymbol{H}]\boldsymbol{A}_\alpha|\Psi\rangle = -\omega_n\tilde{\boldsymbol{\mathcal{O}}}_{\omega_n,m}\boldsymbol{A}_\alpha|\Psi\rangle. \tag{7.8}$$

Note that this means that $\tilde{\boldsymbol{\mathcal{O}}}_{\omega_n,m}$ has positive energy. It is possible to check that (7.8) implies certain corrections to (7.7) at $\mathrm{O}(\frac{1}{\mathcal{N}})$.

It is easy to check that (7.8) is equivalent to

$$\tilde{\boldsymbol{\mathcal{O}}}_{\omega_n,m}\boldsymbol{A}_\alpha\boldsymbol{H}|\Psi\rangle = \boldsymbol{A}_\alpha e^{-\frac{\beta\omega_n}{2}}\boldsymbol{\mathcal{O}}^\dagger_{\omega_n,m}\boldsymbol{H}|\Psi\rangle. \tag{7.9}$$

This equation is equivalent to (7.7) when $|\Psi\rangle$ is an energy eigenstate satisfying $H|\Psi\rangle = E|\Psi\rangle$. In other situations $H|\Psi\rangle$ is an independent descendant and (7.9) gives an independent set of constraints on the definition of $\tilde{\boldsymbol{\mathcal{O}}}_{\omega_n,m}$.

We pause to make a slightly subtle point related to a discussion in [12]. The operator product expansion in the CFT implies that the stress tensor always appears in the one-pion exchange of two local generalized free fields. The Hamiltonian is the zero mode of the stress tensor. Nevertheless, it is consistent for the mirrors to effectively commute with the modes of these operators, but not with the Hamiltonian. This is because if we attempt to express the CFT Hamiltonian in terms of the modes of the GFFs we expect to get an expression involving not just quadratic but also higher order terms.

$$\boldsymbol{H} \doteq \sum_n \omega_n \boldsymbol{a}^\dagger_{\omega_n,m}\boldsymbol{a}_{\omega_n,m} + \cdots + \mathrm{O}(\frac{1}{\mathcal{N}}), \tag{7.10}$$

where the … are similar quadratic terms from other fields and the $\mathrm{O}(\frac{1}{\mathcal{N}})$ terms can be obtained from bulk interactions. As usual, the $\doteq$ in the equation above indicates that this holds within low point correlators. The form of (7.10) is dictated by bulk effective field theory, but a similar expression arises from a careful analysis of boundary correlators.

Now, due to the cutoffs on the set $\mathcal{A}$ above, there is no strict relation between $\boldsymbol{H}$ and other elements $\boldsymbol{A}_\alpha \in \mathcal{A}$. Therefore it is *mathematically consistent* to define the mirrors to have a zero commutator to very high order with ordinary operators but have a nonzero commutator with the Hamiltonian.

However, we must mention another physical point. The $\tilde{\boldsymbol{\mathcal{O}}}_{\omega_n,m}$ operators that we have defined above are auxiliary variables, which do not have any direct physical significance. This is because there is no left asymptotic region in the geometry. It is the $\tilde{\boldsymbol{a}}_{\omega_n,m}$ operators that appear in right-relational observables. Since these observables are defined relationally, they are not strictly local. Therefore, depending on the precise choice of gauge, it is possible—without any loss of locality in the bulk—to consider operators that have a nonzero commutator with $\boldsymbol{a}_{\omega_n,m}$ at subleading $\mathrm{O}(\frac{1}{\mathcal{N}})$. This may even be convenient from some perspectives. We comment more on this issue in forthcoming work.

We now return to the definition of the mirror operators. Equations (7.7) can be considered to be *linear equations* that define the operator $\tilde{\boldsymbol{\mathcal{O}}}_{\omega_n,m}$. We now explain why these equations are consistent.

First, note that if $\boldsymbol{A}_p \in \mathcal{A}_{\rm gff}$ then, in general, we cannot annihilate an equilibrium state by its action,

$$\boldsymbol{A}_p|\Psi\rangle \neq 0, \quad \forall\, \boldsymbol{A}_p \in \mathcal{A}_{\rm gff}. \tag{7.11}$$

This is simply a consequence of the fact that $\dim(\mathcal{A}_{\rm gff}) \ll e^{\mathcal{N}}$ and therefore the space of states annihilated by an element of $\mathcal{A}_{\rm gff}$ is of a very high codimension.

For physical reasons we consider energy eigenstates, which can be annihilated by elements of $\mathcal{A}_{\boldsymbol{H}}$. In such cases, we might have $(\boldsymbol{H}-E)|\Psi\rangle = 0$ for some eigenvalue E. However, as we noted above, in such cases (7.9) reduces to (7.7), and therefore does not lead to an inconsistency.[20]

To summarize (7.7) and (7.9) specify the action of the mirror operator, $\tilde{\mathcal{O}}_{\omega_n,m}$, on a set of linearly independent vectors. This guarantees that we can find a linear operator with the desired action. We can even write down an explicit solution for these linear equations as follows.

We consider a basis of $\mathcal{H}_\Psi$ given by

$$\boldsymbol{A}_1|\Psi\rangle \ldots \boldsymbol{A}_{\mathcal{D}_\mathcal{A}}|\Psi\rangle,$$

and denote an element of this basis by $|v_p\rangle$, where the corresponding $\boldsymbol{A}_p$ may include $\boldsymbol{H}$. The linear equations (7.7) and (7.9) specify the action of the operator $\tilde{\mathcal{O}}_{\omega_n,m}$ on this basis as

$$\tilde{\mathcal{O}}_{\omega_n,m}|v_p\rangle = |u_p\rangle,$$

where $|u_p\rangle$ can be read off from the right-hand side of (7.7) and (7.9). With $g_{pq} = \langle v_p|v_q\rangle$, we can simply define

$$\tilde{\mathcal{O}}_{\omega_n,m} = \sum_{p,q} g^{pq}|u_q\rangle\langle v_p|, \tag{7.12}$$

where g^{pq} is the inverse of g_{pq}. The solution (7.12) has the property that it acts only within $\mathcal{H}_\Psi$. If $\boldsymbol{P}_{\mathcal{H}_\Psi}|w\rangle = 0$ for a state $|w\rangle$, then $\tilde{\mathcal{O}}_{\omega_n,m}|w\rangle = 0$.

This definition directly extends to near-equilibrium states. Given a state of the form (7.6), we define the action of the mirrors by

$$\tilde{\mathcal{O}}_{\omega_n,m}\boldsymbol{A}_\alpha|\Psi^{\rm ne}\rangle = e^{-\frac{\beta\omega_n}{2}}\boldsymbol{A}_\alpha \boldsymbol{U}\mathcal{O}^\dagger_{\omega_n,m}\boldsymbol{U}^{-1}|\Psi^{\rm ne}\rangle. \tag{7.13}$$

The commutator with the Hamiltonian is unchanged.

$$\tilde{\mathcal{O}}_{\omega_n,m}\boldsymbol{H}\boldsymbol{A}_\alpha|\Psi^{\rm ne}\rangle = \boldsymbol{H}\tilde{\mathcal{O}}_{\omega_n,m}\boldsymbol{A}_\alpha|\Psi^{\rm ne}\rangle - \omega_n\tilde{\mathcal{O}}_{\omega_n,m}\boldsymbol{A}_\alpha|\Psi^{\rm ne}\rangle,$$

where all elements on the right-hand side can be computed using (7.13).

[20]Here we have been careful to consider these special states where some descendants obtained by the action of conserved charges are null. In the rest of the paper, when we consider the action of the mirror operators in other settings, we do not always consider this case separately. However, our construction can smoothly accommodate charge or energy eigenstates in all cases.

D. Resolution of paradoxes

We emphasize that our construction above resolves *all* of the paradoxes set out by AMPSS in [2–4]. We reviewed and sharpened these paradoxes in Sec. V but none of these arguments apply to state-dependent operators.

Our construction resolves the $N_a \neq 0$ argument as follows. It is true that typical energy eigenstates are smooth, whereas number eigenstates may not be smooth. However, as we saw in (5.2) to obtain a contradiction we have to perform a *basis change* to go from (5.3) where the trace is evaluated in the energy eigenbasis to (5.5) where the trace is evaluated in the number eigenbasis. If the operator $\boldsymbol{P}_F$ that appears there is state dependent, then this change of basis is impermissible because it is a different operator in each eigenstate. We can see this immediately if we make the state-dependence explicit by adding a small superscript

$$\frac{1}{\mathcal{D}_E}\sum_{\mathcal{R}_E}\langle E|\boldsymbol{P}_F^{\{E\}}|E\rangle \neq \frac{1}{\mathcal{D}_E}\sum_{\mathcal{R}_E}\langle N_i|\boldsymbol{P}_F^{\{N_i\}}|N_i\rangle,$$

even if these two sets of eigenstates span the same space $\mathcal{R}_E$.

In (5.3) we refined the original "lack of a left inverse paradox" of [3] to argue that no state-independent operator could have the commutator required of $\tilde{\boldsymbol{a}}_{\omega_n,m}$ with its adjoint and with the CFT Hamiltonian. However, the argument breaks down if we attempt to apply it to state-dependent operators. In (5.12) we had to use the cyclicity of the trace. But if the operator $\tilde{\boldsymbol{a}}_{\omega_n,m}$ that appears varies as we vary the energy eigenstate then we cannot use this.

As we explained in Sec. V D, the commutator argument is not really a paradox but more of a "genericity argument." Our construction sidesteps this because our mirrors are designed to explicitly commute with the ordinary operators within correlation functions as (7.7) shows.

Finally, consider the strong-subadditivity paradox of [1,2]. Our construction resolves this through a version of black hole complementarity [45,54]. The statement is that it is *impossible* to define mirror operators so that they exactly commute with all CFT operators in any finite time band. From the CFT this is clear from general principles of local quantum field theory. Therefore the mirror operators that describe the interior of the black hole must appear to commute with simple observables within correlation functions but cannot do so exactly. This is a precise version of the colloquial statement that the "interior is a scrambled version of the exterior." The strong-subadditivity paradox assumes that the Hilbert space of gravity factorizes exactly into parts that can be associated with the outside and inside of the black hole. If complementarity is correct, then this assumption is wrong and the strong-subadditivity paradox vanishes.

Note that this resolution to the strong-subadditivity paradox also implies that for some questions—in particular for bulk correlation functions involving $O(\mathcal{N})$ insertions—the notion of locality breaks down completely in the bulk.

This is consistent with the widely held belief that locality is not exact in quantum gravity. However, it is also consistent with various other arguments that suggest a breakdown of locality at this order. For example, it is not difficult to estimate that when one considers correlators with $O(\mathcal{N})$ insertions on the boundary, the $\frac{1}{\mathcal{N}}$ expansion breaks down [55]. Since bulk locality is generally considered to be synonymous with the $\frac{1}{\mathcal{N}}$ expansion, this indicates that bulk locality breaks down at this order. Alternately, from a consideration of scattering amplitudes in bulk effective field theory, it is not difficult to show directly that bulk perturbation theory breaks down when the number of insertions becomes very large [56]. It is possible to see these nonlocal effects even about empty AdS, as we describe in forthcoming work [57]. This nonlocality clearly indicates a rather unusual and profound property of quantum gravity, which deserves further attention.

We direct the reader to [8,9] for further discussion of the resolution of these paradoxes.

E. Small superpositions of equilibrium and near-equilibrium states

We now describe how our construction extends to small superpositions of states. Such superpositions are important, and obtain a direct observational significance, when we consider entangled states of the CFT with an external system of qubits in Sec. IX E. For now we are interested in the following abstract question.

Question: Is exciting a superposition of states by a mirror operator the same as superposing the excited states.

We show that the answer to this question is affirmative. This follows almost trivially from the definition above and ensures that the infalling observer does not observe any departures from linearity.

1. Superpositions of equilibrium states

Consider a superposition of equilibrium states $|\Psi_k\rangle$,

$$|\Psi_s\rangle = \sum_{k=1}^{M} |\Psi_k\rangle, \tag{7.14}$$

where M is an O(1) number and we assume that $\langle\Psi_k|\Psi_p\rangle = 0$ for $k \neq p$ and also that $\sum_k |\langle\Psi_k|\Psi_k\rangle|^2 = 1$ so that the state (7.14) is normalized.

We first show that for generic $|\Psi_k\rangle$, the superposition (7.14) is also in equilibrium. Let us assume that each equilibrium state can be expanded $|\Psi_k\rangle = \sum_i \alpha_{k,i}|E_i\rangle$, so that the entire superposition is

$$|\Psi_s\rangle = \sum_{i,k} \alpha_{k,i}|E_i\rangle.$$

We now consider $\boldsymbol{A}_p \in \mathcal{A}$ and assume that it obeys the eigenstate thermalization hypothesis [43].

$$\langle E_i|\boldsymbol{A}_p|E_j\rangle = A(E_i)\delta_{ij} + e^{-\frac{1}{2}S(\frac{E_i+E_j}{2})}B(E_i,E_j)R_{ij}. \tag{7.15}$$

Here, the quantity $S(\frac{E_i+E_j}{2})$ is the log of the density of states at the mean energy, for which we just write S. The functions A, B are "smooth" functions, and R_{ij} is a matrix with erratically varying phases in its entries but with magnitudes of order 1.

We see now that

$$\langle\Psi_s|\boldsymbol{A}_p|\Psi_s\rangle = \sum_{i,k,n} \alpha^*_{k,i}\alpha_{n,i}A(E_i) + \sum_{i\neq j,k,n} e^{-\frac{1}{2}S(\frac{E_i+E_j}{2})}B(E_i,E_j)R_{ij}\alpha^*_{k,i}\alpha_{n,j}.$$

Consider the first term in the sum above. This involves a sum over $O(e^S)$ energy eigenstates, but for $k \neq n$ the terms in this sum are erratic. Since each $\alpha_{k,i} = O(e^{-\frac{S}{2}})$, this turns into an erratic sum over e^S terms over size e^{-S}. We expect it to typically be only of size $O(e^{-\frac{S}{2}})$. The same argument applies to the second term in the sum, involving R. This term, irrespective of whether $n = k$ or $n \neq k$, turns into an erratic sum over e^{2S} terms, each of size $e^{-\frac{3S}{2}}$. This is again expected to typically only be of size $e^{-\frac{S}{2}}$. This leads to the conclusion that

$$\langle\Psi_s|\boldsymbol{A}_p|\Psi_s\rangle = \sum_{k=1}^{M}\langle\Psi_k|\boldsymbol{A}_p|\Psi_k\rangle + O(e^{-\frac{S}{2}}).$$

Therefore if the equilibrium criterion (7.2) applies to each state $|\Psi_k\rangle$ it also applies to the superposition $|\Psi_s\rangle$, as long as $M = O(1)$. Therefore the superposition is also in equilibrium.

The interesting case is where the $|\Psi_i\rangle$ are microstates corresponding to the same black hole.[21] We can now define the mirrors independently for $|\Psi_s\rangle$ and each of the $|\Psi_i\rangle$. We display this state-dependence explicitly with a superscript below.

We now notice the following simple fact:

$$\tilde{\mathcal{O}}^{\{\mathrm{sup}\}}_{\omega_n,m}\boldsymbol{A}_\alpha|\Psi_s\rangle = e^{-\frac{\beta\omega_n}{2}}\boldsymbol{A}_\alpha\mathcal{O}^\dagger_{\omega_n,m}|\Psi_s\rangle.$$

This follows because $|\Psi_s\rangle$ is also in equilibrium and at the temperature β^{-1}. On the other hand

$$\tilde{\mathcal{O}}^{\{k\}}_{\omega_n,m}\boldsymbol{A}_\alpha|\Psi_k\rangle = e^{-\frac{\beta\omega_n}{2}}\boldsymbol{A}_\alpha\mathcal{O}^\dagger_{\omega_n,m}|\Psi_k\rangle.$$

[21]The case where they correspond to different geometries simply leads to a classical probability distribution over the various possibilities as we described around (3.5). This situation is not of significant physical interest but, in any case, it can be dealt with easily by extending the results obtained here.

Therefore we find that

$$\tilde{\mathcal{O}}^{\{\mathrm{sup}\}}_{\omega_n,m}\boldsymbol{A}_\alpha|\Psi_s\rangle=\sum_{k=1}^{M}\tilde{\mathcal{O}}^{\{k\}}_{\omega_n,m}\boldsymbol{A}_\alpha|\Psi_k\rangle.$$

This equation shows that the mirror operators act consistently with the superposition principle, as long as we are looking at small superpositions of equilibrium states. As we see later, this is important in order for the infalling observer not to be able to detect any violations of quantum mechanics.

2. Superpositions of near-equilibrium states

Now, we consider an O(1) superposition of near-equilibrium states

$$|\Psi^{\mathrm{ne}}_s\rangle=\sum_{k=1}^{M}\boldsymbol{U}_k|\Psi_k\rangle, \tag{7.16}$$

where $|\Psi_k\rangle$ are orthogonal equilibrium states, as previously, and we again assume that the sum in (7.16) is normalized to 1. Here, as in (7.6), $\boldsymbol{U}_k=e^{i\boldsymbol{A}_k}$, where $\boldsymbol{A}_k$ are Hermitian elements of $\mathcal{A}_{\mathrm{gff}}$.

We now define the action of the tildes via

$$\tilde{\mathcal{O}}_{\omega_n,m}\boldsymbol{A}_\alpha|\Psi^{\mathrm{ne}}_s\rangle=\sum_{k=1}^{M}\boldsymbol{A}_\alpha\boldsymbol{U}_k e^{-\frac{\beta\omega_n}{2}}\mathcal{O}^\dagger_{\omega_n,m}|\Psi_k\rangle. \tag{7.17}$$

Note that, strictly speaking, (7.17) is an extension of our definition of mirror operators since a superposition of near-equilibrium states is not itself a near-equilibrium state by the definition of such states in (7.6).

We also note that in this case the action of $\tilde{\mathcal{O}}_{\omega_n,m}$ is not closed within the span of $\mathcal{A}|\Psi^{\mathrm{ne}}_s\rangle$. This can be seen from (7.17) where the right-hand side is not just an ordinary operator acting on $|\Psi^{\mathrm{ne}}_s\rangle$. It is convenient to imagine that we expand the little Hilbert space to the direct sum of the little Hilbert spaces produced by acting on the equilibrium states in (7.16),

$$\mathcal{H}_{\Psi^{\mathrm{ne}}_s}=\bigoplus_k\mathcal{H}_{\Psi_k}.$$

This may be used as a general rule when the space obtained by acting with $\mathcal{A}$ does not contain any equilibrium state at all.

Let us check that (7.17) immediately passes a consistency check. The decomposition of a state in the form (7.16) is not unique. As we explained above, almost all sums of O(1) equilibrium states are also equilibrium states. Correspondingly $\mathcal{H}_{\Psi^{\mathrm{ne}}_s}$ contains many equilibrium states.

This implies that we can just as well write (7.16) as

$$|\Psi^{\mathrm{ne}}_s\rangle=\sum_{k,q,p=1}^{M}\boldsymbol{U}_kQ^{-1}_{kq}Q_{qp}|\Psi_p\rangle=\sum_{q=1}^{M}\boldsymbol{V}_q|\Psi'_q\rangle,$$

with

$$\boldsymbol{V}_q=\sum_{k=1}^{M}\boldsymbol{U}_kQ^{-1}_{kq};\qquad|\Psi'_q\rangle=\sum_{p=1}^{M}Q_{qp}|\Psi_p\rangle.$$

Here Q is any invertible $M\times M$ matrix and Q^{-1} is its inverse: $\sum_q Q^{-1}_{kq}Q_{qp}=\delta_{kp}$. It is important to us that the matrices $\boldsymbol{V}_q$ also be invertible. This is true for generic choices of the $\boldsymbol{U}_k$ and we only consider cases of this sort.

Now, since the state $|\Psi'_q\rangle$ will also typically be in equilibrium, it is equally natural to demand that

$$\tilde{\mathcal{O}}_{\omega_n,m}\boldsymbol{A}_\alpha|\Psi^{\mathrm{ne}}_s\rangle=e^{-\frac{\beta\omega_n}{2}}\boldsymbol{A}_\alpha\sum_{q=1}^{M}\boldsymbol{V}_q\mathcal{O}^\dagger_{\omega_n,m}|\Psi'_q\rangle. \tag{7.18}$$

We ensure that (7.18) is consistent with (7.17). But this follows immediately by inserting the definitions of V and $|\Psi'_q\rangle$ above.

We can also repeat the check we performed for equilibrium states above. Using the definition (7.17) of mirror operators on superpositions of near-equilibrium states on the left-hand side of the equation below, we have

$$\tilde{\mathcal{O}}^{\{\Psi^{\mathrm{ne}}_s\}}_{\omega_n,m}|\Psi^{\mathrm{ne}}_s\rangle=\sum_{k=1}^{M}\tilde{\mathcal{O}}^{\{k\}}_{\omega_n,m}\boldsymbol{A}_\alpha\boldsymbol{U}_k|\Psi_k\rangle, \tag{7.19}$$

where on the right-hand side we use the standard definition of the mirrors on nonequilibrium states given in (7.13), and we have again indicated the state-dependence explicitly by means of the superscript.

The result (7.19) shows that the infalling observer does not observe any violation of linearity even for superpositions of near-equilibrium states. This includes, as a special case, a superposition of an equilibrium and a near-equilibrium state, and thereby answers a question about superposition raised in [58].

F. The interior of the eternal black hole

We conclude this section by constructing state-dependent local operators in the eternal black hole. We already showed in (6.4) that the naive state-independent construction of local operators where we identify $\tilde{\mathcal{O}}_{\omega_n,m}=\mathcal{O}_{L\omega_n,m}$ does not work correctly in the states $|\Psi_{\mathrm{T}}\rangle$ defined in (6.12).

We proceed as follows. We start by reviewing the conditions that we need from the mirrors in the eternal black hole. Based on these, we guess an appropriate solution. We then verify that it meets the conditions that

we outlined. We hasten to add that the formulas we present here can be derived in a completely systematic fashion using the formalism for entangled states that we present in Sec. IX. We present this alternate method of obtaining the answer only because it provides some additional insight into the nature of state-dependence.

We suggest that the reader also consult [23]—where this result is stated concisely—before examining the detailed calculation below.

Constraints on $\tilde{\mathcal{O}}_{\omega_n,m}$: The precise conditions that $\tilde{\mathcal{O}}_{\omega_n,m}$ need to satisfy are given in Sec. VI. These modes need to be correctly entangled with $\mathcal{O}_{\omega_n,m}$ in all states $|\Psi_{\rm T}\rangle$; they need to commute with the $\mathcal{O}_{\omega_n,m}$ within correlators, and also have the commutator with the Hamiltonians given in (6.17).

In fact all of these conditions would be met if

$$\langle\Psi_{\rm T}|\boldsymbol{A}_\alpha\tilde{\mathcal{O}}_{\omega_n,m}\boldsymbol{A}_\beta|\Psi_{\rm T}\rangle = \langle\Psi_{\rm T}|\boldsymbol{A}_\alpha\mathcal{O}_{L\omega_n,m}(T)\boldsymbol{A}_\beta|\Psi_{\rm T}\rangle + {\rm O}\left(\frac{1}{\mathcal{N}}\right), \tag{7.20}$$

where

$$\mathcal{O}_{L\omega_n,m}(T) \equiv \frac{1}{T_{\rm b}^{\frac{1}{2}}}\int_{-T_{\rm b}}^{T_{\rm b}} \mathcal{O}_L(t+T,\Omega)e^{i\omega_n t}Y_m^*(\Omega)dtd^{d-1}\Omega. \tag{7.21}$$

Note that for small T we have $\mathcal{O}_{L\omega_n,m}(T) = \mathcal{O}_{L\omega_n,m}e^{-i\omega T}$. However, this is no longer true when $T \gg T_{\rm b}$. Since we allow exponentially large T in the states $|\Psi_{\rm T}\rangle$, we must adopt the more careful definition (7.21).

We can try and achieve (7.20) through the use of projectors as in Sec. IV B 2. In particular, we use a projector to detect the state as an excitation of $|\Psi_{\rm T}\rangle$ and then modulate $\tilde{\mathcal{O}}_{\omega_n,m}$ accordingly. We caution the reader that this program is only partly successful. But to this end, we investigate these projectors in some detail below. We have to construct these projectors and then in order to put them together correctly, we also need to examine their overlaps.

Projectors on $\mathcal{H}_{\Psi_{\rm T}}$: We define the projector $\mathcal{H}_{\Psi_{\rm T}}$ as follows:

$$\boldsymbol{P}_{\mathcal{H}_{\Psi_{\rm T}}}\boldsymbol{A}_\alpha|\Psi_{\rm T}\rangle = \boldsymbol{A}_\alpha|\Psi_{\rm T}\rangle,$$
$$\text{if} \quad \forall\ \mathbf{A}_\alpha, \qquad \langle v|\boldsymbol{A}_\alpha|\Psi_{\rm T}\rangle = 0 \Rightarrow \boldsymbol{P}_{\mathcal{H}_{\Psi_{\rm T}}}|v\rangle = 0.$$

In these equations we restrict $\boldsymbol{A}_\alpha \in \mathcal{A}_{\rm gff}$ and do not allow it to include $\boldsymbol{H}$.

We can construct the projector explicitly. Define

$$g_{\alpha\beta} = \langle\Psi_{\rm T}|\boldsymbol{A}_\beta^\dagger\boldsymbol{A}_\alpha|\Psi_{\rm T}\rangle.$$

Note that $g_{\alpha\beta}$ is actually independent of T because the operators above come from the right CFT and commute with the left Hamiltonian that is used to evolve $|\Psi_{\rm tfd}\rangle$ to $|\Psi_{\rm T}\rangle$. Then the projector can be written as

$$\boldsymbol{P}_{\mathcal{H}_{\Psi_{\rm T}}} = \sum_{\alpha\beta} g^{\alpha\beta}\boldsymbol{A}_\alpha|\Psi_{\rm T}\rangle\langle\Psi_{\rm T}|\boldsymbol{A}_\beta^\dagger,$$

where $g^{\alpha\beta}$ is the inverse of $g_{\alpha\beta}$. We can check that

$$\boldsymbol{P}_{\mathcal{H}_{\Psi_{\rm T}}}\boldsymbol{A}_\gamma|\Psi_{\rm T}\rangle = \sum_{\alpha\beta} g^{\alpha\beta}\boldsymbol{A}_\alpha|\Psi_{\rm T}\rangle g_{\beta\gamma} = \boldsymbol{A}_\gamma|\Psi_{\rm T}\rangle.$$

Obviously, in the orthogonal subspace, $\boldsymbol{P}_{\mathcal{H}_{\Psi_{\rm T}}}$ gives 0.

Overlaps of the projectors $\boldsymbol{P}_{\mathcal{H}_{\Psi_{\rm T}}}$: Next we have to account for the fact that the different projectors $\boldsymbol{P}_{\mathcal{H}_{\Psi_{\rm T}}}$ are not quite orthogonal for different values of T. We can calculate the overlap between the states $|\Psi_{\rm T}\rangle$ and their descendants as follows. We have

$$\langle\Psi_{\rm tfd}|\boldsymbol{A}_\alpha|\Psi_{\rm T}\rangle = \frac{1}{Z(\beta)}\sum_E e^{-\beta E}\langle E|\boldsymbol{A}_\alpha|E\rangle e^{iET}, \tag{7.22}$$

where all cross terms have dropped out because the operator $\boldsymbol{A}_\alpha$ acts only within the right CFT, and we can use the eigenstates in the left CFT to impose a delta function in energy.

First, let us consider this quantity for $T \ll 1$. In this situation we can approximate (7.22) by

$$\langle\Psi_{\rm tfd}|\boldsymbol{A}_\alpha|\Psi_{\rm T}\rangle = \frac{1}{Z(\beta)}\int e^{-\beta E}e^{S(E)}A(E)e^{iET},$$

where we have indicated the diagonal element of $\boldsymbol{A}_\alpha$ by $A(E)$ as in (7.15).

We can compute this integral using a saddle-point approximation. We write the exponent as

$$-\beta E + S(E) = -\beta E_0 + S(E_0) + \frac{1}{2}(E-E_0)^2\frac{\partial^2 S}{\partial^2 E}\bigg|_{E=E_0},$$

where E_0 satisfies

$$\frac{\partial S}{\partial E}\bigg|_{E=E_0} = \beta.$$

Consider the second derivative term. We write the temperature as a function of energy $\tau(E)$, and then this is just

$$\frac{\partial\frac{1}{\tau(E)}}{\partial E} = -\frac{1}{\tau^2(E)}\partial\frac{\tau(E)}{\partial E} = -\frac{1}{\tau^2(E)C},$$

where C is the specific heat. Note that $C \propto \mathcal{N}$. Evaluated at $E = E_0$, we find

$$\frac{\partial^2 S}{\partial^2 E}\bigg|_{E=E_0} = -\frac{\beta^2}{C}.$$

Therefore the integral above can be written

$$\frac{1}{Z(\beta)}\int \exp\left[-\frac{\beta^2}{C}\frac{(E-E_0)^2}{2}+iET\right]A(E)dE.$$

Now notice that if $A(E)$ is a smooth function of $\frac{E}{\mathcal{N}}$ it varies slowly over the energy scales $\sqrt{C}$ that are relevant here, since $\frac{E}{\mathcal{N}}$ changes only by $\frac{1}{\sqrt{C}}$ over this scale. Second, since we have assumed that $T \ll 1$, we conclude that

$$\langle\Psi_{\rm tfd}|\boldsymbol{A}_\alpha|\Psi_{\rm T}\rangle = \left(\langle\boldsymbol{A}_\alpha\rangle + {\rm O}\left(\frac{1}{\mathcal{N}}\right)\right)e^{-\frac{CT^2}{2\beta^2}}e^{iE_0T}, \quad (7.23)$$

where the expectation value on the right is the normal expectation value taken in $|\Psi_{\rm tfd}\rangle$. Note that we can actually get the prefactor right, and it precisely cancels the factor of $\frac{1}{Z(\beta)}$ in the integral. In particular note that (7.23) also has the correct limit at $T=0$. Below, we use

$$f(T) = e^{-\frac{CT^2}{2\beta^2}}e^{iE_0T}.$$

We caution the reader that the estimates for the overlap between different projectors are no longer valid for $T \sim {\rm O}(1)$. We consider this case separately below.

Guess for $\tilde{\mathcal{O}}_{\omega_n,m}$: We can now use these projectors and the idea explained above to write down a guess for the $\tilde{\mathcal{O}}_{\omega_n,m}$ that will reproduce (7.20). We consider

$$\tilde{\mathcal{O}}_{\omega_n m} = \sqrt{\frac{C}{\pi\beta^2}}\int_{-T_{\rm cut}}^{T_{\rm cut}} \mathcal{O}_{L\omega_n m}(T_i)\boldsymbol{P}_{\mathcal{H}_{\Psi_{T_i}}}dT_i, \quad (7.24)$$

where $T_{\rm cut}$ is a cutoff that we explore further below. The idea of (7.24) is that the projector $\boldsymbol{P}_{\mathcal{H}_{\Psi_{T_i}}}$ detects the state it is acting on as an excitation of $|\Psi_{T_i}\rangle$, and therefore the insertion of $\tilde{\mathcal{O}}_{\omega_n,m}$ effectively turns into an insertion of $\mathcal{O}_{L\omega_n m}(T_i)$ as required in (7.20).

We now verify in detail that the guess (7.24) does satisfy all the conditions that we need in the state $|\Psi_{\rm tfd}\rangle$ and in states $|\Psi_{\rm T}\rangle$ for $|T| < T_{\rm cut}$. For states where T does not satisfy this condition we need to change the operator (7.24) as we describe below.

Correlators of $\tilde{\mathcal{O}}_{\omega_n m}$: We are interested in inserting the proposed mirror defined in (7.24) in correlators. We find that

$$\begin{aligned}&\langle\Psi_{\rm T}|\boldsymbol{A}_\alpha\tilde{\mathcal{O}}_{\omega_n m}\boldsymbol{A}_\beta|\Psi_{\rm T}\rangle\\ &= \sqrt{\frac{C}{\pi\beta^2}}\int_{-T_{\rm cut}}^{T_{\rm cut}}dT_i\\ &\quad\times\langle\Psi_{\rm tfd}|e^{-i\boldsymbol{H}T}\boldsymbol{A}_\alpha\mathcal{O}_{L\omega_n m}(T_i)\boldsymbol{P}_{\mathcal{H}_{\Psi_{T_i}}}\boldsymbol{A}_\beta e^{i\boldsymbol{H}T}|\Psi_{\rm tfd}\rangle.\end{aligned}$$

To evaluate the integral on the right-hand side we consider the integrand

$$\begin{aligned}&\langle\Psi_{\rm tfd}|e^{-i\boldsymbol{H}T}\boldsymbol{A}_\alpha\boldsymbol{P}_{\mathcal{H}_{\Psi_{T_i}}}\boldsymbol{A}_\beta e^{i\boldsymbol{H}T}|\Psi_{\rm tfd}\rangle\\ &= \langle\Psi_{\rm tfd}|\boldsymbol{A}_\alpha\boldsymbol{P}_{\mathcal{H}_{\Psi_{T-T_i}}}\boldsymbol{A}_\beta|\Psi_{\rm tfd}\rangle\\ &= \sum_{\gamma\delta}\langle\Psi_{\rm tfd}|\boldsymbol{A}_\alpha g^{\gamma\delta}\boldsymbol{A}_\gamma|\Psi_{T-T_i}\rangle\langle\Psi_{T-T_i}|\boldsymbol{A}_\delta^\dagger\boldsymbol{A}_\beta|\Psi_{\rm tfd}\rangle,\end{aligned}$$

where we have first used the factors of $e^{i\boldsymbol{H}T}$ to convert the projector to $\boldsymbol{P}_{\mathcal{H}_{\Psi_{T-T_i}}}$ and then we have inserted the explicit expression for the projector derived above. This quantity can be further be simplified to

$$\begin{aligned}&\langle\Psi_{\rm tfd}|\boldsymbol{A}_\alpha\boldsymbol{P}_{\mathcal{H}_{\Psi_{T-T_i}}}\boldsymbol{A}_\beta|\Psi_{\rm tfd}\rangle\\ &= |f(T-T_i)|^2\sum_{\gamma\delta}\langle\Psi_{\rm tfd}|\boldsymbol{A}_\alpha g^{\gamma\delta}\boldsymbol{A}_\gamma|\Psi_{\rm tfd}\rangle\langle\Psi_{\rm tfd}|\boldsymbol{A}_\delta^\dagger\boldsymbol{A}_\beta|\Psi_{\rm tfd}\rangle\\ &= |f(T-T_i)|^2\langle\Psi_{\rm tfd}|\boldsymbol{A}_\alpha\boldsymbol{P}_{\mathcal{H}_{\Psi_{\rm tfd}}}\boldsymbol{A}_\beta|\Psi_{\rm tfd}\rangle\\ &= |f(T-T_i)|^2\langle\Psi_{\rm tfd}|\boldsymbol{A}_\alpha\boldsymbol{A}_\beta|\Psi_{\rm tfd}\rangle,\end{aligned}$$

where we have used the expression for mixed correlators in (7.23), then reabsorbed the sum over γ, δ into another projector, and recognized that the projector acts as the identity on descendants of $|\Psi_{\rm tfd}\rangle$.

Plugging this into the original integral we find that

$$\begin{aligned}&\langle\Psi_{\rm T}|\boldsymbol{A}_\alpha\tilde{\mathcal{O}}_{\omega_n m}\boldsymbol{A}_\beta|\Psi_{\rm T}\rangle\\ &= \sqrt{\frac{C}{\pi\beta^2}}\int_{-T_{\rm cut}}^{T_{\rm cut}}dT_i|f(T-T_i)|^2\\ &\quad\times\langle\Psi_{\rm tfd}|e^{-i\boldsymbol{H}T}\boldsymbol{A}_\alpha\mathcal{O}_{L\omega_n m}(T_i)\boldsymbol{A}_\beta e^{i\boldsymbol{H}T}|\Psi_{\rm tfd}\rangle\\ &= \langle\Psi_{\rm T}|\boldsymbol{A}_\alpha\mathcal{O}_{L\omega_n m}(T)\boldsymbol{A}_\beta|\Psi_{\rm T}\rangle + {\rm O}\left(\frac{1}{\mathcal{N}}\right).\end{aligned}$$

Here we have used the fact that $\mathcal{O}_{L\omega_n m}(T_i)$ varies very slowly with respect to the function $f(T-T_i)$, provided $\omega_n \ll \mathcal{N}$ since $C \sim {\rm O}(\mathcal{N})$. Therefore, to leading order in $\frac{1}{\mathcal{N}}$ we can simply evaluate this integral in the saddle-point approximation which leads to the result above. This result is, of course, valid provided that $|T| < T_{\rm cut}$ and it agrees with what was required in (7.20).

Note that this immediately leads to the right two-point and higher point functions. For example,

$$\begin{aligned}\langle\Psi_{\rm T}|\mathcal{O}_{L\omega_n m}(T)\mathcal{O}_{\omega_n m}|\Psi_{\rm T}\rangle &= \langle\Psi_{\rm tfd}|\mathcal{O}_{L\omega_n m}\mathcal{O}_{\omega_n m}|\Psi_{\rm tfd}\rangle\\ &= e^{\frac{-\beta\omega_n}{2}}G_\beta(\omega_n,m),\end{aligned}$$

which is precisely what is required.

Commutator with Hamiltonians: Finally we check the behavior of the proposed $\tilde{\mathcal{O}}_{\omega_n m}$ under time evolution with the left and right Hamiltonians. Notice that

$$\boldsymbol{P}_{\mathcal{H}_{\Psi_{T_i}}}e^{-i\boldsymbol{H}T} = e^{-i\boldsymbol{H}T}\boldsymbol{P}_{\mathcal{H}_{\Psi_{T_i+T}}}.$$

Therefore,

$$e^{iHT}\tilde{\mathcal{O}}_{\omega_n m}e^{-iHT} = \sqrt{\frac{C}{\pi\beta^2}}\int_{-T_{\rm cut}}^{T_{\rm cut}} \mathcal{O}_{L\omega_n m}(T_i)\boldsymbol{P}_{\mathcal{H}_{\Psi_{T_i+T}}} dT_i$$
$$= \sqrt{\frac{C}{\pi\beta^2}}\int_{T-T_{\rm cut}}^{T_{\rm cut}+T} \mathcal{O}_{L\omega_n m}(T_i - T)\boldsymbol{P}_{\mathcal{H}_{\Psi_{T_i}}} dT_i,$$

where the last equality comes from a change of variables inside the integral. Note that

$$\mathcal{O}_{L\omega_n m}(T_i - T) = e^{i\omega_n T}\mathcal{O}_{L\omega_n m},$$

for $T \sim \mathrm{O}(1)$, and as long as $T \ll T_{\rm b}$. Now, when inserted into correlation functions, the cutoffs are exponentially irrelevant as the analysis above shows. The dominant contribution when $\tilde{\mathcal{O}}_{\omega_n m}$ is inserted into a correlator always comes from a saddle point in the interior of the integral. Therefore we find that within correlation functions

$$e^{iHT}\tilde{\mathcal{O}}_{\omega_n m}e^{-iHT} \doteq e^{i\omega_n T}\tilde{\mathcal{O}}_{\omega_n m},$$

which is precisely what is required as long as we do not evolve for a very long time.

A very similar analysis shows that conjugation by $e^{i\boldsymbol{H}_L T}$ leaves $\tilde{\mathcal{O}}_{\omega_n m}$ invariant within correlators because of the transformation of $\mathcal{O}_{L\omega_n m}$ in the integral above. This completes our verification of (6.17).

1. Analysis of state-dependence in the eternal black hole

The reader should note that our construction is explicitly state dependent. The operators (7.24) fail to click correctly when they are inserted in states $|\Psi_{\rm T}\rangle$ with $T \gg T_{\rm cut}$. It is easy to verify this by repeating the exercise above. The reader will find that when $\tilde{\mathcal{O}}_{\omega_n,m}$ is inserted into a correlator, the saddle point of the integral over T_i occurs outside the range of integration, and therefore the correlator is exponentially suppressed.

Now, we might naively believe that this can be fixed simply by taking $T_{\rm cut}$ to infinity. However, we show below that if we do this, then instead of behaving correctly in every state, the integral (7.24) would fail to behave correctly in any state. To see this we need to reconsider the overlap estimate of (7.23). The expression in (7.23) is not the correct answer for $T \gg 1$ since our saddle-point technique of evaluating the thermal correlator breaks down if the phase factor that arises from the term involving T varies too rapidly.

At large T, we simply note that the overlap is a sum over approximately $\mathrm{O}(e^S)$ uncorrelated complex numbers of $\mathrm{O}(1)$.

$$\langle\Psi_{\rm tfd}|\boldsymbol{A}_\alpha|\Psi_{\rm T}\rangle = \frac{1}{Z(\beta)}\sum e^{-\beta E}e^{iET}A(E) = \mathrm{O}(e^{-\frac{S}{2}}),$$
$$T \gg 1. \tag{7.25}$$

In particular for $T \gg 1$, this overlap is much larger than the overlap predicted by (7.23). It has a fat tail.

Therefore if we take $T_{\rm cut}$ to be exponentially large, $T_{\rm cut} \gg \mathrm{O}(e^S)$, and insert (7.24) into a correlator, then the contributions from this fat tail will overwhelm the contribution of the dominant saddle. This is the reason that we are forced to use state-dependence.

For the states $|\Psi_{\rm T}\rangle$ with $T \gg e^S$, we still write down interior operators. These operators are given by

$$\mathcal{O}^{\{T\}}_{\omega_n,m} = \sqrt{\frac{C}{\pi\beta^2}}\int_{T-T_{\rm cut}}^{T+T_{\rm cut}} \mathcal{O}_{L\omega_n m}(T_i)\boldsymbol{P}_{\mathcal{H}_{\Psi_{T_i}}} dT_i,$$

where we have explicitly moved the range of integration.

This discussion helps to shed light on the nature of state-dependence. By performing these large diffeomorphisms we have, in a sense, "geometrized" the microstates of the black hole. The states $|\Psi_{\rm T}\rangle$ are all identical states from the perspective of the right-infalling observer, but the left and right modes are entangled differently in each of them. The novel part of this situation is that these are also distinct and well-separated solutions from the point of view of the semiclassical theory if we keep track of how the solution is glued to the boundary.

Now, classically the right-relational observables are well-defined objects on each of these geometries. Often, in such situations, it is possible to lift such classical observables to operators as we describe in more detail in Appendix A. This is usually done by identifying classical solutions as coherent states in the Hilbert space, and using projectors to map classical functions to operators. [See, for instance, (A4).] However, if we consider the states $|\Psi_{\rm T}\rangle$ for exponentially large ranges of T, then (7.25) tells us they are "overcomplete." This overcompleteness goes beyond the usual overcompleteness of coherent states. In fact, we believe that a computation using coherent states to represent the different states $|\Psi_{\rm T}\rangle$ in canonical gravity should yield the overlap (7.23) but at large T this is very different from (7.25). This forces us to use state-dependent operators for the black hole interior, even in this one-parameter class of states.

By considering time-shifted versions of the geon solution analyzed in [42], we believe that it should not be difficult to find a similar one-parameter set in a single CFT where state-dependence can be analyzed in detail.

VIII. REMOVING AMBIGUITIES IN THE CONSTRUCTION

We now turn to the issue of some ambiguities in our construction. There are two sorts of ambiguities that have

been described in the literature. The first is related to an observation about the eternal black hole by Marolf and Wall [47] and a similar observation by van Raamsdonk [11] which was framed more directly in terms of our construction. We show here how this ambiguity should be resolved.

The second ambiguity was discussed by the authors of [4] and some of these objections were expanded in a paper by Harlow [12]. However, Harlow's construction attempted to add to this ambiguity by adopting a modified definition of the mirror operators, which had a different commutator with the Hamiltonian from the one in our construction. We show that this alternate definition of the mirror operators of [12] suffers from certain inconsistencies which we point out below.

As a consequence of this, the alternate mirror operators described by Harlow do not themselves have direct physical significance. However, it is true that there is an interesting class of excited states that we consider in Sec. VIII C; these are related to the analysis of [12] but we consider them independently so as to separate them from the main claims of that paper.

We should mention that an additional class of ambiguities, involving only ordinary operators, was described in [3]. The authors of [3] suggested that one could act with the Schwarzschild number operator $e^{i\theta N_\omega}|\Psi\rangle$ on an equilibrium state to obtain another state that was approximately time-translationally invariant. We have addressed this issue previously. (See page 46 of [9].) If we use a finite time band to extract the modes of the CFT generalized free fields, and then combine them into a number operator then such an operator does not commute exactly with the CFT Hamiltonian. One may attempt to improve this construction by considering an extremely slow acting source, which inserts only a finite amount of energy into the system over an extremely long time scale. The action of such a source might be consistent with our equilibrium condition but this would not be a contradiction since the infalling observer would also not see any excitation in this case.

A. Mirror unitary behind the horizon

Consider an equilibrium state $|\Psi\rangle$ and perform the construction described in Sec. VII, leading to the mirror operators. Now, consider the state

$$|\Psi_{\rm ex}\rangle = e^{i\alpha\tilde{\boldsymbol{A}}_p}|\Psi\rangle \equiv \tilde{\boldsymbol{U}}|\Psi\rangle. \tag{8.1}$$

Here $\tilde{\boldsymbol{A}}_p$ is the mirror of a Hermitian operator satisfying $(\boldsymbol{A}_p)^\dagger = \boldsymbol{A}_p$. The parameter α is a real number that is useful below.

In our construction above, we have not really defined the exponentiated version of the mirror operators. To exponentiate the mirror we need to be able to evaluate

$$e^{i\alpha\tilde{\boldsymbol{A}}_p}|\Psi\rangle = \sum_{n=0}^{\infty}\frac{(i\alpha)^n}{n!}(\tilde{\boldsymbol{A}}_p)^n|\Psi\rangle,$$

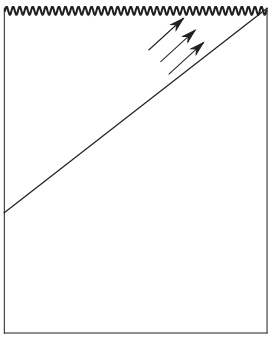


FIG. 11. A state $|\Psi_{\rm ex}\rangle = \tilde{\boldsymbol{U}}|\Psi\rangle$ corresponding to an equilibrium state $|\Psi\rangle$ excited with a mirror unitary behind the horizon $\tilde{\boldsymbol{U}}$.

which involves arbitrarily high products of the mirror operator and necessarily takes us outside the space $\mathcal{H}_\Psi$. To be precise, beyond some cutoff K, we expect $\langle\Psi|[(\tilde{\boldsymbol{A}}_p)^K, \boldsymbol{A}_s]|\Psi\rangle \neq 0$. The precise value of K depends on the precise definition of $\tilde{\boldsymbol{A}}_p$. We return to this edge effect below in the discussion of Harlow's ambiguity.

The first putative ambiguity mentioned in the beginning of Sec. VIII is the following: if we assume that the state $|\Psi\rangle$ is a black hole in an equilibrium state, then the state $|\Psi_{\rm ex}\rangle$ should be an excited state. Intuitively we expect $|\Psi_{\rm ex}\rangle$ to be a state with an excitation behind the horizon as shown in Fig. 11. In particular, an observer crossing the horizon in the state $|\Psi_{\rm ex}\rangle$, within a suitable time range, should detect this excitation. Now, the question is, suppose we are given the state $|\Psi_{\rm ex}\rangle$ without the additional information that it came by acting with $e^{i\alpha\tilde{\boldsymbol{A}}_p}$ on some equilibrium state $|\Psi\rangle$. How can we directly detect that the state $|\Psi_{\rm ex}\rangle$ is a nonequilibrium state? The difficulty comes from the fact that since $\tilde{\boldsymbol{U}}$ approximately commutes with elements of the small algebra, we have

$$\begin{aligned}&\langle\Psi|\tilde{\boldsymbol{U}}^\dagger\boldsymbol{\mathcal{O}}_{\omega_1,m_1}\ldots\boldsymbol{\mathcal{O}}_{\omega_n,m_n}\tilde{\boldsymbol{U}}|\Psi\rangle\\ &\quad= \langle\Psi|\boldsymbol{\mathcal{O}}_{\omega_1,m_1}\ldots\boldsymbol{\mathcal{O}}_{\omega_n,m_n}|\Psi\rangle + R,\end{aligned}$$

where R is the small remainder that we discussed above. We neglect this remainder in what follows. Hence, simple correlators of the small algebra on the state $|\Psi_{\rm ex}\rangle$ seem to be almost the same as those in the state $|\Psi\rangle$. This might lead to the erroneous conclusion that $|\Psi_{\rm ex}\rangle$ is an equilibrium state. This mistake would lead to the definition of mirror operators as if $|\Psi_{\rm ex}\rangle$ were equilibrium, and using these wrong mirror operators would lead to the incorrect prediction that the infalling observer will not detect any excitation behind the horizon. In order to avoid this ambiguity in the mirror operator construction, we need to find a way to detect from the CFT that $|\Psi_{\rm ex}\rangle$ is an excited state.

The key point is that we have also included the Hamiltonian in our set of observables. The Hamiltonian does *not* commute with the mirror operators. Hence, correlators of operators in the small algebra, together with insertions of the Hamiltonian, differ between typical

equilibrium states and states which have been excited by mirror unitary operators $|\Psi_{ex}\rangle = \tilde{U}|\Psi\rangle$. We can use these differences as a diagnostic of the nonequilibrium nature of these states. This resolves the ambiguity of the mirror unitaries behind the horizon.

To make this more clear, let us consider the state $|\Psi_{ex}\rangle$ in (8.1) and let us define

$$\tilde{\boldsymbol{A}}_s \equiv [\boldsymbol{H}, \tilde{\boldsymbol{A}}_p]. \tag{8.2}$$

We can detect the nonequilibrium nature of the state $|\Psi_{ex}\rangle$ by considering the correlation function with $\boldsymbol{H}$ and the corresponding $\boldsymbol{A}_s$ operator

$$\begin{aligned}
&\langle\Psi_{ex}|\boldsymbol{H}\boldsymbol{A}_s|\Psi_{ex}\rangle \\
&\quad = \langle\Psi|\tilde{\boldsymbol{U}}^\dagger \boldsymbol{H}\boldsymbol{A}_s\tilde{\boldsymbol{U}}|\Psi\rangle \\
&\quad = \langle\Psi|(\mathbf{1} - i\alpha\tilde{\boldsymbol{A}}_p)\boldsymbol{H}\boldsymbol{A}_s(\mathbf{1} + i\alpha\tilde{\boldsymbol{A}}_p)|\Psi\rangle + \mathrm{O}(\alpha^2) \\
&\quad = \langle\Psi|\boldsymbol{H}\boldsymbol{A}_s|\Psi\rangle + i\alpha\langle\Psi|\tilde{\boldsymbol{A}}_s\boldsymbol{A}_s|\Psi\rangle + \mathrm{O}(\alpha^2) \\
&\quad = \mathrm{O}(e^{-\frac{S}{2}}) + i\alpha\langle\Psi|\boldsymbol{A}_s e^{-\frac{\beta\boldsymbol{H}}{2}}(\boldsymbol{A}_s)^\dagger e^{\frac{\beta\boldsymbol{H}}{2}}|\Psi\rangle + \mathrm{O}(\alpha^2).
\end{aligned} \tag{8.3}$$

Here we have used the fact the equilibrium expectation value of the operator $\boldsymbol{H}\boldsymbol{A}_s$ is exponentially small, if $\boldsymbol{A}_s$ has nonzero energy. On the other hand, we expect that the expectation value in the second term of the last line above is O(1). So, we see that for the observable in (8.3), we discern a substantial deviation from its equilibrium value. This allows us to classify the state $|\Psi_{ex}\rangle$ as an "excited state," as expected intuitively.

For a concrete example, let us take $\tilde{\boldsymbol{A}}_p$ in (8.1) to be $\tilde{\boldsymbol{A}}_p = \tilde{\mathcal{O}}_{\omega,m} + \tilde{\mathcal{O}}^\dagger_{\omega,m}$.[22] We consider (8.2) for this case, to find $\tilde{\boldsymbol{A}}_s = \omega(\tilde{\mathcal{O}}_{\omega,m} - \tilde{\mathcal{O}}^\dagger_{\omega,m})$. In an equilibrium state we have

$$\omega\langle\Psi|\boldsymbol{H}(\mathcal{O}_{\omega,m} - \mathcal{O}^\dagger_{\omega,m})|\Psi\rangle = 0, \tag{8.4}$$

up to exponentially small corrections. On the other hand, for the state $e^{i\alpha(\tilde{\mathcal{O}}_{\omega,m}+\tilde{\mathcal{O}}^\dagger_{\omega,m})}|\Psi\rangle$ we find to linear order in α and up to exponentially small corrections that

$$\begin{aligned}
&\omega\langle\Psi|\tilde{\boldsymbol{U}}^\dagger\boldsymbol{H}(\mathcal{O}_{\omega,m} - \mathcal{O}^\dagger_{\omega,m})\tilde{\boldsymbol{U}}|\Psi\rangle \\
&= \omega\langle\Psi|e^{-i\alpha(\tilde{\mathcal{O}}_{\omega,m}+(\tilde{\mathcal{O}}_{\omega,m})^\dagger)}\boldsymbol{H}(\mathcal{O}_{\omega,m} - \mathcal{O}^\dagger_{\omega,m})e^{i\alpha(\tilde{\mathcal{O}}_{\omega,m}+\tilde{\mathcal{O}}^\dagger_{\omega,m})}|\Psi\rangle \\
&= i\alpha\omega^2\langle\Psi|(\tilde{\mathcal{O}}_{\omega,m} - \tilde{\mathcal{O}}^\dagger_{\omega,m})(\mathcal{O}_{\omega,m} - \mathcal{O}^\dagger_{\omega,m})|\Psi\rangle + \mathrm{O}(\alpha^2) \\
&= i\alpha\omega^2\langle\Psi|(\mathcal{O}_{\omega,m} - \mathcal{O}^\dagger_{\omega,m})\left(e^{-\frac{\beta\omega}{2}}\mathcal{O}^\dagger_{\omega,m} - e^{\frac{\beta\omega}{2}}\mathcal{O}_{\omega,m}\right)|\Psi\rangle \\
&\quad + \mathrm{O}(\alpha^2) \\
&= 2i\alpha\omega^2 e^{-\frac{\beta\omega}{2}}G_\beta(\omega, m) + \mathrm{O}(\alpha^2),
\end{aligned}$$

which is O(1). So this correlator is different on $|\Psi_{ex}\rangle$ from that on the equilibrium state (8.4) and by measuring this correlator we can detect the excitation by the mirror unitary behind the horizon.

Uniqueness of the behind-horizon unitaries: We note that given a state $|\Psi_{ex}\rangle$ of the form (8.1) it has an essentially unique decomposition into an equilibrium state and a unitary behind the horizon. The reason is as follows. First, it is clear that we cannot have such a decomposition with two different basis states, since in that case we would have

$$\tilde{\boldsymbol{U}}_{\mathbf{1}}|\Psi_1\rangle = \tilde{\boldsymbol{U}}_{\mathbf{2}}|\Psi_2\rangle \Rightarrow |\Psi_1\rangle = \tilde{\boldsymbol{U}}^\dagger_{\mathbf{1}}\tilde{\boldsymbol{U}}_{\mathbf{2}}|\Psi_2\rangle.$$

As we have shown above, if $|\Psi_2\rangle$ is in equilibrium a relation of the sort above implies that $|\Psi_1\rangle$ cannot be in equilibrium, and vice versa.

Furthermore, with $\tilde{\boldsymbol{U}}_{\mathbf{1}} = e^{i\tilde{\boldsymbol{A}}_1}$, and $\tilde{\boldsymbol{U}}_{\mathbf{2}} = e^{i\tilde{\boldsymbol{A}}_2}$, it is clear from a chain of reasoning that

$$\begin{aligned}
\tilde{\boldsymbol{U}}_{\mathbf{1}}|\Psi\rangle = \tilde{\boldsymbol{U}}_{\mathbf{2}}|\Psi\rangle &\Rightarrow (\tilde{\boldsymbol{U}}^\dagger_{\mathbf{1}}\tilde{\boldsymbol{U}}_{\mathbf{2}})|\Psi\rangle = |\Psi\rangle \Rightarrow (\tilde{\boldsymbol{A}}_1 - \tilde{\boldsymbol{A}}_2)|\Psi\rangle = 0 \\
&\Rightarrow (\boldsymbol{A}^\dagger_1 - \boldsymbol{A}^\dagger_2)|\Psi\rangle = 0,
\end{aligned}$$

which is prohibited by (7.11) unless $\boldsymbol{A}_1 = \boldsymbol{A}_2$, and so $\tilde{\boldsymbol{U}}_{\mathbf{1}} = \tilde{\boldsymbol{U}}_{\mathbf{2}}$. This concludes our proof of the uniqueness of the decomposition.

Therefore, to summarize, given a state of the form (8.1) we can not only detect that it is out of equilibrium, but even detect the operator with which it has been excited.

B. Comments on the Harlow unitaries

Now, let us turn to a second set of unitaries described by Harlow [12], who attempted to define a new set of mirror operators $\tilde{\boldsymbol{X}}^H_{\omega,m}$ which act on an equilibrium state as follows:

$$\tilde{\boldsymbol{X}}^H_{\omega,m}\boldsymbol{A}_\beta|\Psi\rangle = \boldsymbol{A}_\beta e^{-\frac{\beta\omega}{2}}(\mathcal{O}_{\omega,m})^\dagger|\Psi\rangle, \tag{8.5}$$

$$[\tilde{\boldsymbol{X}}^H_{\omega,m}, \boldsymbol{H}]\boldsymbol{A}_\beta|\Psi\rangle \stackrel{?}{=} 0. \tag{8.6}$$

Notice that the first equation, (8.5), is the same as the one in our definition, (7.7), but the commutator with the Hamiltonian given in (8.6) differs from ours, which is specified by (7.8).

We now show that the definition of mirror operators given by Harlow is inconsistent, and runs into difficulties in several physical situations. We discuss an energy eigenstate, and then a state drawn from the microcanonical ensemble.[23] We then discuss a more serious problem—definition (8.6) leads to operators that do not satisfy the Heisenberg equations of motion. Therefore, these operators

[22] In this section and in Sec. IX, to lighten the notation, instead of ω_n for the discretized frequencies, we drop the subscripts and simply write ω.

[23] This was already noticed in [12] and discussed in Sec. II D of that paper.

$\tilde{\boldsymbol{X}}^H_{\omega,m}$ cannot be used to build up gauge-invariant relational observables.

1. Inconsistency of $\tilde{X}^H_{\omega,m}$ mirrors in energy eigenstates

First, we point out that the second line above, (8.6), does not have any solutions at all, when defined about energy eigenstates. We find that

$$\tilde{\boldsymbol{X}}^H_{\omega,m}\boldsymbol{H}|E\rangle = E\tilde{\boldsymbol{X}}^H_{\omega,m}|E\rangle = e^{-\frac{\beta\omega}{2}}E\boldsymbol{\mathcal{O}}^\dagger_{\omega,m}|E\rangle. \tag{8.7}$$

But[24]

$$\begin{aligned}\boldsymbol{H}\tilde{\boldsymbol{X}}^H_{\omega,m}|E\rangle &\overset{?}{=} e^{-\frac{\beta\omega}{2}}\boldsymbol{H}(\boldsymbol{\mathcal{O}}_{\omega,m})^\dagger|E\rangle \\ &= e^{-\frac{\beta\omega}{2}}[\boldsymbol{H},(\boldsymbol{\mathcal{O}}_{\omega,m})^\dagger]|E\rangle + e^{-\frac{\beta\omega}{2}}(\boldsymbol{\mathcal{O}}_{\omega,m})^\dagger\boldsymbol{H}|E\rangle \\ &= e^{-\frac{\beta\omega}{2}}\omega\boldsymbol{\mathcal{O}}^\dagger_{\omega,m}|E\rangle + e^{-\frac{\beta\omega}{2}}E\boldsymbol{\mathcal{O}}^\dagger_{\omega,m}|E\rangle \\ &= e^{-\frac{\beta\omega}{2}}(E+\omega)\boldsymbol{\mathcal{O}}^\dagger_{\omega,m}|E\rangle. \end{aligned}\tag{8.8}$$

To understand the inconsistency of Harlow's definition for eigenstates, we consider the correlator $\langle E|\boldsymbol{\mathcal{O}}_{\omega,m}[\boldsymbol{H},\tilde{\boldsymbol{X}}^H_{\omega,m}]|E\rangle$. We can compute it in two ways. The first is to subtract (8.7) from (8.8) and multiply the resulting state with the bra $\langle E|\boldsymbol{\mathcal{O}}_{\omega,m}$. This leads to the prediction

$$\begin{aligned}\langle E|\boldsymbol{\mathcal{O}}_{\omega,m}[\boldsymbol{H},\tilde{\boldsymbol{X}}^H_{\omega,m}]|E\rangle &= e^{-\frac{\beta\omega}{2}}\langle E|\boldsymbol{\mathcal{O}}_{\omega,m}\omega\boldsymbol{\mathcal{O}}^\dagger_{\omega,m}|E\rangle \\ &= \omega e^{-\frac{\beta\omega}{2}}G_\beta(\omega,m).\end{aligned}\tag{8.9}$$

On the other hand, using directly (8.6), we find that

$$\langle E|\boldsymbol{\mathcal{O}}_{\omega,m}[\boldsymbol{H},\tilde{\boldsymbol{X}}^H_{\omega,m}]|E\rangle \overset{?}{=} 0. \tag{8.10}$$

Clearly (8.10) and (8.9) are in contradiction, and therefore Eq. (8.6), which was used by Harlow to define the mirrors, is actually inconsistent in an energy eigenstate. Moreover note that at this level the contradiction arises at O(1) and cannot be resolved by $\frac{1}{\mathcal{N}}$ corrections.

Now, we move away from a strict energy eigenstate and turn to a state with an O(1) spread in energies. We show that even in such a state, the modified definition of the mirror operators in [12] cannot be used consistently.

2. Inconsistency of $\tilde{\mathbf{X}}^H_{\omega,m}$ in microcanonical states

We now show that the inconsistency in Harlow's unitaries is not restricted to energy eigenstates. It persists in states that are drawn from a microcanonical ensemble with an O(1) spread in energies. Consider a state of the following kind,

$$|\Psi_{\rm mic}\rangle = \sum_i \alpha_i|E_i\rangle,$$

where the coefficients α_i have the property that they are peaked around a given energy, which we call E, but the spread in energies is O(1). More precisely, we demand

$$\begin{aligned}\langle\Psi_{\rm mic}|\boldsymbol{H}|\Psi_{\rm mic}\rangle &= E, \\ \langle\Psi_{\rm mic}|\boldsymbol{P}_E|\Psi_{\rm mic}\rangle &= 1-\mathrm{O}(\mathcal{N}^{-1}),\end{aligned}$$

where

$$\boldsymbol{P}_E = \sum_{i=E-\Delta}^{i=E+\Delta}|E_i\rangle\langle E_i| \tag{8.11}$$

is the projector onto states in the range $E\pm\Delta$, and $\Delta \ll \mathcal{N}$ is some O(1) number.

Now, the key point is as follows. In (8.6) we have imposed the relation that the commutator of the operator $\tilde{\boldsymbol{X}}^H$ with the Hamiltonian annihilates the state. However, the projector onto a range of energies, like the one that appears in (8.11), is also a good observable. In fact, physically we expect to be able to measure this observable rather easily both on the boundary and in the bulk. On the boundary, this observable is completely determined by considering the zero mode of the stress tensor. In the bulk, it can be determined by considering the subleading falloff in the metric. This is in contrast to a projector onto a Schwarzschild number eigenstate which, as we reviewed in Appendix C of [9], requires an extremely long time to measure and projects the final state onto a firewall.

Now, consider again the relation (8.6), but extended to products of the operator $\tilde{\boldsymbol{X}}^H_{\omega,m}$. As we discussed above, unless we can define such products consistently to a high order, it is not possible to consider unitaries made out of this operator, which are required to produce the ambiguity that was discussed in [12].

However, for any O(1) frequency ω, we have an O(1) number n_c, so that

$$n_c\omega > 2\Delta.$$

Now, following (8.6), we impose

$$(\tilde{\boldsymbol{X}}^H_{\omega,m})^{n_c}|\Psi_{\rm mic}\rangle = e^{-\frac{n_c\beta\omega}{2}}(\boldsymbol{\mathcal{O}}^\dagger_{\omega,m})^{n_c}|\Psi_{\rm mic}\rangle + \frac{1}{\mathcal{N}}|\mathcal{R}^{\rm micro}_C\rangle,$$

where we have included a small possible $\frac{1}{\mathcal{N}}$ correction with the property that

$$\langle\mathcal{R}^{\rm micro}_C|\mathcal{R}^{\rm micro}_C\rangle = \mathrm{O}(1).$$

[24]Note that these results are unaffected by a possible small correction to the commutator between the Hamiltonian and the ordinary operator: $\boldsymbol{\mathcal{R}}_C = [\boldsymbol{H},\boldsymbol{\mathcal{O}}^\dagger_{\omega,m}] - \omega\boldsymbol{\mathcal{O}}^\dagger_{\omega,m}$. This may arise because we define the modes by considering only a finite-time interval as we discussed above. However, we expect that $\|\boldsymbol{\mathcal{R}}_C|E\rangle\|^2 \ll 1$, and particularly that $\langle E|\boldsymbol{\mathcal{O}}_{\omega,m}\boldsymbol{\mathcal{R}}_C|E\rangle = \mathrm{O}(\frac{1}{\mathcal{N}})$. These statements just point out that the remainder is small and, in particular, it does not have an overlap with $\boldsymbol{\mathcal{O}}^\dagger_{\omega,m}|E\rangle$ at O(1).

However, now we note that

$$e^{-n_c\beta\omega}\langle\Psi_{\rm mic}|(\mathcal{O}_{\omega,m})^{n_c}\boldsymbol{P}_E(\mathcal{O}^\dagger_{\omega,m})^{n_c}|\Psi_{\rm mic}\rangle \ll 1. \quad (8.12)$$

This is because the action of n_c insertions of $\tilde{\mathcal{O}}^\dagger_{\omega,m}$ raises the energy by the state by $n_c\omega$ and so necessarily takes it out of the band $E\pm\Delta$. On the other hand, if the operator $\tilde{\boldsymbol{X}}^H_{\omega,m}$ is defined to commute also with $\boldsymbol{P}_E$ then we would expect

$$\begin{aligned}&\langle\Psi_{\rm mic}|[(\tilde{\boldsymbol{X}}^H_{\omega,m})^\dagger]^{n_c}\boldsymbol{P}_E(\tilde{\boldsymbol{X}}^H_{\omega,m})^{n_c}|\Psi_{\rm mic}\rangle\\ &\overset{?}{=}\langle\Psi_{\rm mic}|\boldsymbol{P}_E[[(\tilde{\boldsymbol{X}}^H_{\omega,m})^\dagger]^{n_c}(\tilde{\boldsymbol{X}}^H_{\omega,m})^{n_c}|\Psi_{\rm mic}\rangle+\mathrm{O}(\mathcal{N}^{-1}).\\ &=\langle\Psi_{\rm mic}|\boldsymbol{P}_E(\mathcal{O}^\dagger_{\omega,m})^{n_c}(\mathcal{O}_{\omega,m})^{n_c}|\Psi_{\rm mic}\rangle+\mathrm{O}(\mathcal{N}^{-1})\\ &=\mathrm{O}(1),\end{aligned} \quad (8.13)$$

where in the final result we have noted that action of $(\mathcal{O}_{\omega,m})^{n_c}$ followed by the action of its adjoint maps us back to the same band of energies. Clearly the results of (8.12)–(8.13) are in contradiction given the general results about the expectation value of projectors in states that are almost parallel, which we reviewed in Sec. V A.

3. Failure of $\tilde{X}^H_{\omega,m}$ to satisfy the Heisenberg equations of motion

Now we turn to an even more serious difficulty with the mirror operators defined by (8.6): their failure to satisfy the Heisenberg equations of motion. This failure persists even in states with a canonical spread of energies. In such states, the fundamental relation (8.6) does not suffer from an obvious inconsistency, unlike in energy eigenstates or states with a microcanonical spread. However, as we show below these operators nevertheless do not have the correct geometric properties to play the role of interior mirror operators.

In particular, as we described in detail in Sec. VI C 1, if the bulk operators are defined relationally with respect to the boundary, in order to be gauge invariant, then they must satisfy

$$e^{i\boldsymbol{H}T}\boldsymbol{\phi}(t,r,\Omega)e^{-i\boldsymbol{H}T}=\boldsymbol{\phi}(t+T,r,\Omega).$$

It is clear that if we attempt to create these operators by means of the operators defined in (8.6), then the local operators will not obey the Heisenberg equations of motion. Let us check this explicitly by computing a two-point function across the horizon.

Outside the horizon we have the usual expansion of the field in terms of CFT modes,

$$\begin{aligned}\boldsymbol{\phi}^H(t,r_*,\Omega)\underset{U\to0^-}{\to}&\sum_{m,\omega}\frac{1}{\sqrt{\omega C_\beta(\omega,m)}}\\ &\times\mathcal{O}_{\omega,m}e^{-i\omega t}Y_m(\Omega)(e^{i\delta}e^{i\omega r_*}+e^{-i\delta}e^{-i\omega r_*})+\mathrm{H.c.}\end{aligned}$$

This expansion does not depend on our definition of the mirror operators. Inside the horizon, however, using the Harlow mirror operators we find

$$\begin{aligned}\boldsymbol{\phi}^H(t,r_*,\Omega)\underset{U\to0^+}{\to}&\sum_{m,\omega}\frac{e^{-i\delta}}{\sqrt{\omega C_\beta(\omega,m)}}[\mathcal{O}_{\omega,m}e^{-i\omega(t+r_*)}Y_m(\Omega)\\ &+\tilde{\boldsymbol{X}}^H_{\omega,m}e^{i\omega(t-r_*)}Y^*_m(\Omega))+\mathrm{H.c.}].\end{aligned}$$

Now, let us compute correlation functions with this operator in an *equilibrium state*, $|\Psi\rangle$. Moving to the usual Kruskal coordinates U, V, let us consider two points, so that one of them, (U_1,V_1,Ω_1), is just outside the horizon whereas the other (U_2,V_2,Ω_2) is just inside. Then we find

$$\begin{aligned}&\langle\Psi|e^{-i\boldsymbol{H}T}\boldsymbol{\phi}^H(U_1,V_1,\Omega_1)\boldsymbol{\phi}^H(U_2,V_2,\Omega_2)e^{i\boldsymbol{H}T}|\Psi\rangle\\ &=\sum_{m,\omega}\frac{1}{\omega C_\beta(\omega,m)}\\ &\quad\times\left[\langle\mathcal{O}_{\omega,m}\mathcal{O}^\dagger_{\omega,m}\rangle\left(\frac{V_1}{V_2}\right)^{i\omega}+e^{i\omega T}\langle\mathcal{O}_{\omega,m}\tilde{\boldsymbol{X}}^H_{\omega,m}\rangle\left(\frac{-U_1}{U_2}\right)^{i\omega}\right]\\ &\quad\times Y_m(\Omega_1)Y^*_m(\Omega_2)+\mathrm{H.c.}\end{aligned}$$

Notice the extra factor of $e^{i\omega T}$ which appears in front of the $\frac{U_1}{U_2}$ factor. In particular, this implies that if we compute the derivative of the two-point function and take the two points to be close then we find (using the techniques of Sec. IV), substituting the relevant two-point functions and converting the sum to an integral, that

$$\begin{aligned}&\lim_{V_1-V_2\to0}\langle\Psi|e^{-i\boldsymbol{H}T}\partial_U\boldsymbol{\phi}^{\rm H}(U_1,V_1,\Omega_1)\partial_U\boldsymbol{\phi}^{\rm H}\\ &\quad\times(U_2,V_2,\Omega_2)e^{i\boldsymbol{H}T}|\Psi\rangle\\ &=c\frac{\delta^{d-1}(\Omega_1-\Omega_2)}{(U_1-U_2e^{-\frac{2\pi T}{\beta}})^2}.\end{aligned}$$

However, this is in explicit contradiction with the universal short distance form of the correlator that we derived in (4.3). In fact, such a correlator would suggest the presence of a firewall.

Therefore, we have reached the following conclusion. Even in an equilibrium state, where we expect correlation functions to be time invariant, if one uses Harlow's definition of the mirror operators, this leads to the prediction that if one starts with a state with no firewall, a firewall appears immediately.

This is a straightforward consequence of the fact that these putative mirror operators do not obey the Heisenberg equations of motion. The commutator with the Hamiltonian (8.6) was derived neither from a gauge fixing procedure, which we carried out carefully in [9], nor a careful consideration of relational observables in the geometry, which we performed in Sec. VI C.

In fact, the source of this error is apparent. The motivation of [12] to propose the vanishing commutator of the interior operators with the Hamiltonian (8.6) was partly based on the analogy with the thermofield doubled state. In fact, it was argued in [12] that in some specific pure states, one may expect bulk correlators to approximate thermofield correlators to high orders in $\frac{1}{N}$. However, *even* in the thermofield state, as we showed in Sec. VI, when one carefully consider commutators of the right Hamiltonian with the mirrors that are relevant for the right-relational observables, one finds nonzero commutators. It is only if one uses the naive but incorrect expansion of Sec. VI D that one obtains the incorrect expectation for the commutator used in (8.6).

One possible interpretation for $\tilde{\boldsymbol{X}}^H_{\omega,m}$ is that they actually correspond to the operators from the left CFT in a thermofield doubled state and not to the operators behind the horizon at all. This would explain why they do not satisfy the properties expected of the $\tilde{\mathcal{O}}_{\omega,m}$ operators. However, as we showed in Sec. VII F, the subtleties and paradoxes associated with the $\tilde{\mathcal{O}}_{\omega,m}$ construction enter precisely when one attempts to map operators from the left CFT into the operators that a right-infalling observer would see behind the horizon. So, if this alternate interpretation is correct, then the operators $\tilde{\boldsymbol{X}}^H_{\omega,m}$ pertain to a formal construction that is not directly relevant for the construction of the black hole interior in AdS/CFT.

C. States in the "canonical" ensemble

We now turn precisely to an interesting class of excitations of states in the canonical ensemble. The point is that we need to refine our notion of equilibrium, since the time independence of correlators of single-trace operators may not be sufficient to classify these states into equilibrium and nonequilibrium. We do not explicitly perform this classification here, but we show that such a classification should exist.

These states were also discussed in [12], but we phrase the issue independently of Harlow's mirror operators, since these do not have any geometric significance as we pointed out above.

Consider a state $|\Psi_{\rm can}\rangle$ that satisfies the following condition. For any element $\boldsymbol{A}_p$ of the set of observables $\mathcal{A}$, we have

$$\langle\Psi_{\rm can}|\boldsymbol{A}_p|\Psi_{\rm can}\rangle = {\rm Tr}(\boldsymbol{\rho}\boldsymbol{A}_p) + {\rm O}(e^{-S}), \tag{8.14}$$

where $\boldsymbol{\rho}$ is an *invertible* matrix. Note that if the state $|\Psi_{\rm can}\rangle$ is in equilibrium then the density matrix $\boldsymbol{\rho}$ satisfies $[\boldsymbol{H},\boldsymbol{\rho}] = 0$. This is important for correlation functions to be time-translationally invariant.

We pause to make two important points. Given a state $|\Psi_{\rm can}\rangle$ the density matrix that appears on the right of (8.14) is not unique. In fact, the possible solutions to this equation are the subject of entropy maximization [59]. Second, both the energy eigenstate and the sharp microcanonical state that we considered above are not relevant here. We cannot find any *invertible* choice of $\boldsymbol{\rho}$ to satisfy (8.14) for these states without making some matrix elements of the inverse arbitrarily large.

Now, given any Hermitian element of the set of observables $\boldsymbol{A}_p$, we consider the transformation

$$|\Psi'_{\rm can}\rangle = \boldsymbol{\rho}^{\frac{1}{2}} e^{i\boldsymbol{A}_p}\boldsymbol{\rho}^{-\frac{1}{2}}|\Psi_{\rm can}\rangle. \tag{8.15}$$

We can check that correlators of elements of $\mathcal{A}$ in the state $|\Psi'_{\rm can}\rangle$ are the same as those in $|\Psi_{\rm can}\rangle$. We see that

$$\begin{aligned}\langle\Psi'_{\rm can}|\boldsymbol{A}_m|\Psi'_{\rm can}\rangle &= \langle\Psi_{\rm can}|\boldsymbol{\rho}^{-\frac{1}{2}}e^{-i\boldsymbol{A}_p}\boldsymbol{\rho}^{\frac{1}{2}}\boldsymbol{A}_m\boldsymbol{\rho}^{\frac{1}{2}}e^{i\boldsymbol{A}_p}\boldsymbol{\rho}^{-\frac{1}{2}}|\Psi_{\rm can}\rangle \\ &= {\rm Tr}\left[\boldsymbol{\rho}\left(\boldsymbol{\rho}^{-\frac{1}{2}}e^{-i\boldsymbol{A}_p}\boldsymbol{\rho}^{\frac{1}{2}}\boldsymbol{A}_m\boldsymbol{\rho}^{\frac{1}{2}}e^{i\boldsymbol{A}_p}\boldsymbol{\rho}^{-\frac{1}{2}}\right)\right] + {\rm O}(e^{-S}) \qquad (8.16)\\ &= {\rm Tr}(\boldsymbol{\rho}\boldsymbol{A}_m) + {\rm O}(e^{-S}) = \langle\Psi_{\rm can}|\boldsymbol{A}_m|\Psi_{\rm can}\rangle + {\rm O}(e^{-S}). \qquad (8.17)\end{aligned}$$

In obtaining (8.16), we simply used (8.14), and then we use the cyclicity of the trace and (8.14) to obtain the final result in (8.17). The question now is as follows: is the state $|\Psi'_{\rm can}\rangle$ in equilibrium or not?

Consider a concrete example. Take the state that was discussed in [12],

$$|\Psi_{\rm can}\rangle = \frac{1}{\sqrt{Z(\beta)}}\sum_i e^{-\frac{\beta E_i}{2}}e^{i\phi_i}|E_i\rangle, \tag{8.18}$$

where ϕ_i are arbitrary phases, the sum is over all energy eigenstates and $Z(\beta)$ is the partition function of the boundary theory. As discussed in [12] for simple correlators this state behaves like the canonical ensemble to exponential accuracy, and for this state we can take $\boldsymbol{\rho} = \frac{1}{Z(\beta)}e^{-\beta\boldsymbol{H}}$ and satisfy (8.14).

To see this, consider *any* operator, $\boldsymbol{A}_p$, obeying the eigenstate thermalization hypothesis (7.15). Adopting the notation of (7.15), we consider

$$\begin{aligned}\langle\Psi_{\rm can}|\boldsymbol{A}_p|\Psi_{\rm can}\rangle = &\frac{1}{Z(\beta)}\sum_i A(E_i)e^{-\beta E_i} \\ &+\frac{1}{Z(\beta)}\sum_{ij}e^{-\beta\frac{E_i+E_j}{2}}R_{ij}e^{-S}B(E_i,E_j)e^{i(\phi_j-\phi_i)}.\end{aligned}$$

To convert the second term to a sum over i, we sum over all j that can be connected by the cross terms. We make the further reasonable assumption that the unitary links states that are separated only by a finite band, i.e. $B(E_i,E_j) \ll 1$ for $|E_i - E_j| \gg 1$. Now, we see that for each value of i, the sum over j runs over effectively ${\rm O}(e^S)$ states. However, since these states contribute with varying phases the typical size of this sum over j is suppressed by $e^{-\frac{S}{2}}$ compared to the first term involving $A(E_i)$. So we can estimate that

$$\langle\Psi_{\rm can}|\boldsymbol{A}_p|\Psi_{\rm can}\rangle = \frac{1}{Z(\beta)}\sum_i A(E_i)e^{-\beta E_i} + \mathrm{O}(e^{-\frac{S}{2}}) = \frac{1}{Z(\beta)}\mathrm{Tr}(e^{-\beta H}\boldsymbol{A}_p) + \mathrm{O}(e^{-\frac{S}{2}}).$$

Now, we consider the group of transformations of the form (8.15) that we can make to this state, where now $\boldsymbol{\rho} = \frac{1}{Z(\beta)}\mathrm{Tr}(e^{-\beta \boldsymbol{H}})$,

$$\boldsymbol{M}|\Psi_{\rm can}\rangle \equiv e^{-\frac{\beta \boldsymbol{H}}{2}}e^{i\boldsymbol{A}_p}e^{\frac{\beta \boldsymbol{H}}{2}}|\Psi_{\rm can}\rangle. \tag{8.19}$$

The question is, if $|\Psi_{\rm can}\rangle$ is an equilibrium state, then is $\boldsymbol{M}|\Psi_{\rm can}\rangle$ in equilibrium or not?

We work with this concrete example to consider this question. Of course, the reader can easily generalize this discussion to states that mimic a density matrix that is distinct from the thermal one.

At first sight, this question is a little puzzling because of two seemingly contradictory facts. On the one hand, all correlators of elements of $\mathcal{A}$ in this new state (8.19) are the same as in the canonical ensemble

$$\langle\Psi_{\rm can}|\boldsymbol{M}^\dagger\boldsymbol{A}_m\boldsymbol{M}|\Psi_{\rm can}\rangle = \frac{1}{Z(\beta)}\mathrm{Tr}(e^{-\beta H}\boldsymbol{A}_m) + \mathrm{O}(e^{-\frac{S}{2}}).$$

On the other hand, it is easy to verify that

$$\langle\Psi_{\rm can}|\boldsymbol{M}^\dagger e^{-i\tilde{\boldsymbol{A}}_p}|\Psi_{\rm can}\rangle = 1 - \mathrm{O}(\mathcal{N}^{-1}), \tag{8.20}$$

where here $e^{-i\tilde{\boldsymbol{A}}_p}|\Psi_{\rm can}\rangle$ is an excited state, as discussed in Sec. VIII A. So if we declare the transformed state in (8.19) as an equilibrium state, then we would have the unusual situation of having equilibrium and excited states separated by a distance $\frac{1}{\mathcal{N}}$ in the Hilbert space (8.20). This would not be a contradiction, since the operators $\tilde{\mathcal{O}}$ are state dependent, but it would be a rather striking departure from the behavior of state-independent operators.

Therefore, the better alternative is to enlarge the set of observables $\mathcal{A}$ to include an operator that can distinguish between the states $\boldsymbol{M}|\Psi_{\rm can}\rangle$ and $|\Psi_{\rm can}\rangle$. There are many such operators because it is certainly not true that all physical properties of these states can be captured by the thermal density matrix. For example, if we take the boundary to be on S^{d-1} and ask for the entanglement entropy of a subregion on this boundary, then in both states, this entanglement entropy starts to decrease after the volume of the subregion increases past half the volume of the S^{d-1}, which would not be the case for a truly thermal mixed state.

We return to the discussion of the appropriate operators that can detect this excitation in future work. However, for now, we perform an important consistency check. Consider the set of states formed by the action of the group of exponentiated unitaries

$$\{|\Psi_{\rm can}\rangle, \boldsymbol{M}(\boldsymbol{A}_1)|\Psi_{\rm can}\rangle, \boldsymbol{M}(\boldsymbol{A}_2)|\Psi_{\rm can}\rangle \ldots \boldsymbol{M}(\boldsymbol{A}_n)|\Psi_{\rm can}\rangle\}, \tag{8.21}$$

where $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots \boldsymbol{A}_n$ are elements of $\mathcal{A}$ and $\boldsymbol{M}(\boldsymbol{A}_p)|\Psi_{\rm can}\rangle \equiv e^{-\frac{\beta H}{2}}e^{i\boldsymbol{A}_p}e^{\frac{\beta H}{2}}|\Psi_{\rm can}\rangle$ as above. We show that it is consistent, in principle, to have sets of this form, where only one element of the set is an equilibrium state, and all others are nonequilibrium states. The consistency check that we need to perform is to ensure that such a classification will not violate the rule that most states in the Hilbert space must be equilibrium states.

1. Consistency condition for maps from equilibrium to nonequilibrium states

Let us state this consistency condition more precisely. It is applicable not only to this case, but to more general statistical mechanical questions of classifying equilibrium. Let us say that we have two regions of the Hilbert space, $\mathcal{D}$, and $\mathcal{I}$. We have a function on the Hilbert space $\Theta_E(\Psi)$, with the property that $\Theta_E(\Psi) = 0$ for equilibrium states and $\Theta_E(\Psi) = 1$ for nonequilibrium states. This function provides a classification of equilibrium. Next, we have a measure on the Hilbert space $d\mu(\Psi)$, which has the property that by this measure most states in both $\mathcal{D}$ and $\mathcal{I}$ are in equilibrium.

$$\frac{\int_{\mathcal{D}} d\mu(\Psi)\Theta_E(\Psi)}{\int_{\mathcal{D}} d\mu(\Psi)} \ll 1, \tag{8.22}$$

$$\frac{\int_{\mathcal{I}} d\mu(\Psi)\Theta_E(\Psi)}{\int_{\mathcal{I}} d\mu(\Psi)} \ll 1. \tag{8.23}$$

This means that the volume of nonequilibrium states as a fraction of the total volume is very small both in $\mathcal{D}$ and in $\mathcal{I}$. Finally, consider a map M,

$$M\colon\ \mathcal{D} \to \mathcal{I},$$

which has the property that it maps equilibrium to nonequilibrium states.

Let $M(\mathcal{D})$ be the image of $\mathcal{D}$ under this map. Now, let $\mathcal{I}_{\mathcal{D}}$ be the region of the Hilbert space that is within a distance ϵ of the set $\mathcal{I}$. More precisely, for $\epsilon \ll 1$,

$$|\Psi_{\rm ex}\rangle \in \mathcal{I}_{\mathcal{D}} \Leftrightarrow \exists |\Psi\rangle \in \mathcal{D}, \quad \text{s.t.} \quad |\langle\Psi_{\rm ex}|M|\Psi\rangle|^2 \geq 1 - \epsilon^2. \tag{8.24}$$

Then we have the following important *consistency condition* on this map:

$$\frac{\int_{\mathcal{I}_{\mathcal{D}}} d\mu(\Psi)}{\int_{\mathcal{I}} d\mu(\Psi)} \ll 1. \tag{8.25}$$

We explain this condition in a little more detail below. Intuitively, it means that states that are close to the image of $\mathcal{D}$ under M must have very small volume in $\mathcal{I}$.

From this condition it follows immediately that an invertible map $\mathcal{D} \to \mathcal{D}$ *cannot* map equilibrium to non-equilibrium states consistently. For example, consider the microcanonical measure where we pick states in an energy band. (We define this more precisely below.) We expect most such states to be in equilibrium. Now consider time translations, which map this region back to itself. Therefore, the image under time translations of the original region is the region itself. Thus time translations do not satisfy (8.25) and therefore cannot have the property.

2. Microcanonical ensemble and unitaries

To warm up for the problem of maps from canonical states back to themselves, we consider a similar problem for the microcanonical ensemble. We define this ensemble, define an appropriate measure so that (8.22)–(8.23) are satisfied and show how unitaries of simple operators do satisfy (8.25).

Consider the set of all states of the form

$$|\Psi_{\rm mic}\rangle = \sum_{E_i=E-\Delta}^{E_i=E+\Delta} a_i |E_i\rangle, \tag{8.26}$$

where $\sum_i |a_i|^2 = 1$ for the state to be normalized. We now write down an invariant Haar measure on this set, $d\mu(\Psi_{\rm mic})$, with the property that for *any unitary* that maps states of the form (8.26) back to another state of the same form, $|\Psi'_{\rm mic}\rangle = \mathbf{U}|\Psi_{\rm mic}\rangle$, we have $d\mu(\Psi'_{\rm mic}) = d\mu(\Psi_{\rm mic})$. Explicitly, to obtain the microcanonical ensemble, we consider the uniform probability measure

$$d\mu(a_i) = \delta\left(1 - \sum_i |a_i|^2\right) d^2a_1 \ldots d^2a_D, \tag{8.27}$$

where D is the total number of energy eigenstates in this range, and N_μ is a normalization constant that we fix below. In the measure above, note that we have not identified states that differ by a phase.

In terms of the objects introduced in Sec. VIII C 1, the set $\mathcal{D}$ is the set of all states of the form (8.26). We have not specified a precise equilibrium function. However, with almost any reasonable choice of $\Theta_E(\Psi)$, for example, we can choose this function so that it implements our equilibrium condition in (7.5), and with the measure (8.27), we see that (8.22) is satisfied.

We can take the map under consideration to be the unitary matrix, $\boldsymbol{U}_m = e^{i\boldsymbol{A}_m}$. Now one might naively imagine that there are "as many" states of the form $\boldsymbol{U}_m|\Psi_{\rm mic}\rangle$ as of the form $|\Psi_{\rm mic}\rangle$. The reason this is still consistent with the fact that most states are equilibrium states is that $\boldsymbol{U}_m|\Psi_{\rm mic}\rangle$ does *not* belong to the original microcanonical ensemble. Even if we consider $\boldsymbol{A}_m = \boldsymbol{\mathcal{O}}_\omega + \boldsymbol{\mathcal{O}}_\omega^\dagger$ where ω is a very low frequency we see that the new state $\boldsymbol{U}_m|\Psi_{\rm mic}\rangle$ contains energy eigenstates of higher energies. The term $\frac{\boldsymbol{A}_m^k}{k!}$ in the expansion of the unitary operator leads to a new ensemble with states $E \pm \Delta \pm k\omega$. The point is that even a small increase in energy increases the volume of the ensemble by a huge amount, and therefore the state $\boldsymbol{U}(\boldsymbol{A}_m)|\Psi_{\rm mic}\rangle$ come from a larger ensemble, where they are extremely atypical.

Let us see this more precisely; let us define $\mathcal{I}$ to be the set of states that can be written in the form (8.26), but with a width $\Delta' > \Delta$. In the example above, if we take $\frac{\Delta'-\Delta}{\omega} \gg 1$, then we can consistently think of the unitary as a map from $\mathcal{D}$ to $\mathcal{I}$. Strictly speaking the image of the lower dimensional manifold in the higher dimensional manifold is measure 0. However, this does mean that nonequilibrium states are infinitely unlikely. To answer physical questions we must examine how many states in the higher dimensional manifold are within an ϵ distance of the states obtained by exciting the lower dimensional manifold with a unitary. The relevance of this condition is that by the arguments of Sec. V A the expectation value of any projector in states which have an almost unit inner product is almost identical and therefore such states have similar physical properties.

To verify (8.25), we consider the volume of the manifold of all states of the form (8.26). This is just given by integrating the measure (8.27) which results in

$$V_{\rm micro} = \frac{\pi^D}{\Gamma(D)},$$

the factor of (2π) coming from the integral over the phases in each coefficient in the state.

The action of the unitary maps this into a slightly larger ensemble. The larger ensemble has dimension D' and total volume

$$V_{\rm exc} = \frac{\pi^{D'}}{\Gamma(D')}.$$

Now we may consider the volume of the set of states that are within a distance ϵ of the image of the unitary map, in the sense of (8.24). This volume can be calculated through the following integral:

$$\begin{aligned} V_{\rm image} &= \int_0^\epsilon 2x dx \int \delta\left(1 - x^2 - \sum_i |a_i|^2\right) d^2a_1 \ldots d^2a_D \\ &\quad \times \int \delta\left(x^2 - \sum_j |b_j|^2\right) d^2b_1 \ldots d^2b_{D'-D} \\ &= \frac{\pi^{D'}}{\Gamma(D)\Gamma(D'-D)} \int_0^{\epsilon^2} d(x^2)(x^2)^{D'-D-1}(1-x^2)^{D-1}. \end{aligned}$$

The last integral can be represented as an incomplete beta function, but we can bound its value rather easily. First note

that the integrand reaches a maximum at $x^2 = \frac{D'-D-1}{D'-2}$. If ϵ is sufficiently small, then this maximum value is out of the region of integration and the integral is bounded above by

$$V_{\text{image}} < \frac{\pi^{D'}}{\Gamma(D)\Gamma(D'-D)} \epsilon^{2(D'-D)} (1-\epsilon^2)^{D-1}.$$

The ratio of the volume of this region to the volume of the ensemble is given by

$$\frac{V_{\text{image}}}{V_{\text{exc}}} < \frac{\Gamma(D')}{\Gamma(D)\Gamma(D'-D)} \epsilon^{2(D'-D)} (1-\epsilon^2)^{D-1}.$$

In our case, D, D', $D-D'$ are all very large and we can approximate this using Stirling's approximation to obtain

$$\frac{V_{\text{image}}}{V_{\text{exc}}} < (1-\epsilon^2)^{-1} \left(\frac{D'(1-\epsilon^2)}{D}\right)^{D} \left(\frac{D'\epsilon^2}{D'-D}\right)^{D'-D}.$$

In the regime where $\epsilon^2 \ll \frac{D'-D}{D'}$, we see that this ratio is very small.[25]

Therefore even if the unitary increases the dimension of the new ensemble by only a small fraction, it is completely consistent with thermodynamic expectations to classify almost all states both in the original ensemble, and in the new ensemble, as equilibrium states.

3. Excitations of canonical states

Now we want to show that the same principle holds for the canonical states that we discussed above. More precisely, we consider some possible measures on a subset of the Hilbert space, so that typical states picked using this measure are of the form (8.18). Then the action of the operators $\boldsymbol{M}$ takes us to another subset of the Hilbert space where the image of the original subset occupies a vanishingly small volume. By the remark below (8.18), as a corollary, this provides some evidence for the claim that there is no subset of the CFT Hilbert space, with a nice measure satisfying (8.22) which has the property that it is left invariant by the action of $\boldsymbol{M}$.

First, let us attempt to make precise what we mean by states of the form (8.18). In (8.18) we ensured that each coefficient was precisely the Boltzmann factor. This is clearly a very special class of states and we would set ourselves too simple a problem by focusing on these states.

[25]The reader should note that this regime is somewhat different from the regime considered recently in [60]. Indeed, as pointed out there, if we define nearby states by taking $\epsilon^2 \sim \frac{D'-D}{D'}$ then the volume of the image and nearby states is almost the entire volume of the excited manifold. This is not in contradiction with our result above that excited states are atypical. Rather it is the statement that once we move a distance $(\frac{D'-D}{D'})^{\frac{1}{2}}$ from the excited state, we are back in the set of typical states of the Hilbert space.

So we can generalize this slightly to consider states of the form

$$|\Psi_{\text{can}}\rangle = \sum_{E_1}^{E_2} \frac{1}{\sqrt{Z(\beta)}} a_i e^{-\frac{\beta E_i}{2}} |E_i\rangle, \tag{8.28}$$

where the a_i are complex numbers that are drawn from a distribution so that their norms can each independently fluctuate a little about 1 but

$$\langle |a_i|^2 \rangle = 1. \tag{8.29}$$

We comment more on the range of the sum $[E_1, E_2]$ below. It is easy to verify, by repeating the argument above, that even for the states (8.28) we have

$$\langle \Psi_{\text{can}} | \boldsymbol{A}_p | \Psi_{\text{can}} \rangle = \frac{1}{Z(\beta)} \text{Tr}(e^{-\beta H} \boldsymbol{A}_p) + \text{O}(e^{-\frac{S}{2}}).$$

By the central limit theorem, since there is an exponentially large number of energy eigenstates in (8.28), the fact that the coefficients a_i can fluctuate in magnitudes as well as phases is unimportant. To see this consider a range of energies of size $e^{-\frac{\mathcal{N}}{2}}$. Even this tiny range of energies has an exponentially large number of eigenstates. In the notation of (7.15), the expectation value $A(E_i)$ is constant over this range, and therefore the fluctuations of $|a_i|^2$ average out. Therefore, for any smooth function, it is only the mean magnitude of the $|a_i|^2$ that matters, which is what leads to the result above.

Now consider the action of an element of $\boldsymbol{M}$ on the state (8.28). We write $\boldsymbol{M} = e^{-\frac{\beta \mathbf{H}}{2}} \boldsymbol{U} e^{\frac{\beta \mathbf{H}}{2}}$. If the matrix elements of $\boldsymbol{U}$ are $\boldsymbol{U}|E_i\rangle = \sum_j U_{ji} |E_j\rangle$, then we reach the new state

$$\begin{aligned} |\Psi'_{\text{can}}\rangle &\equiv N_M \boldsymbol{M} |\Psi_{\text{can}}\rangle \\ &= N_M \sum_{E_i=E_1}^{E_i=E_2} \sum_{E_j} \frac{1}{\sqrt{Z(\beta)}} e^{-\beta E_j} a_i U_{ji} |E_j\rangle, \end{aligned}$$

where the factor

$$N_M = \langle \Psi_{\text{can}} | \boldsymbol{M}^\dagger \boldsymbol{M} | \Psi_{\text{can}} \rangle^{-\frac{1}{2}}$$

is required to normalize the state. If we neglect the edge effects for the moment (these are important below), then we see that we again have a state of the form (8.28), although with coefficients

$$a_j{}' = N_M \sum_i U_{ij} a_i.$$

From the argument above we can check that $N_m = 1 + \text{O}(e^{-\frac{S}{2}})$. Therefore the action of the group of

transformations denoted by $\boldsymbol{M}$ is basically like that of a unitary transformation on the elements a_i.

We now see the following.

(1) Physically the range of energies that is relevant in (8.28) is limited. So, we may truncate this range so that the lower bound is $E_1 = E - \Delta$ and the upper bound is $E_2 = E + \Delta$. In that case, by an extension of the arguments of the previous subsection we find that $\boldsymbol{M}$ maps us to a slightly larger band of energies. Under almost any reasonable measure, this larger band has a much larger volume and therefore (8.25) is met. The technical details of this argument are identical to the previous subsection since, as we noted, $\boldsymbol{M}$ acts precisely as a unitary transformation on the coefficients a_i.

(2) We may try and avoid this conclusion in the following artificial manner. We extend the band of energies $[E_1, E_2]$ in (8.28) so that it spans a very large range. We now truncate the action of $\boldsymbol{M}$ so that it acts only within this large energy range. By construction, now $\boldsymbol{M}$ maps this set back to itself. This may suggest that (8.25) cannot be met. This conclusion is clearly physically incorrect since the higher energies in (8.28) are physically unimportant and therefore artificially extending the band should have no effect. However, there is another important point. If we indeed take our original domain $\mathcal{D}$ to be the subspace of this large range of energies, and attempt to define a measure that is left invariant by the action of $\boldsymbol{M}$, then as we show below we find that the states (8.28) are extremely unlikely states and themselves occupy only a small volume of the space.

The point is that there is a *tension* between the requirement (8.29) which mandates that all the a_i must have equal and approximately unit magnitude and the fact that $\boldsymbol{M}$ acts as a unitary on this space. We now consider one particular example to bring out this tension. In an attempt to write down a measure that is invariant under the action of $\boldsymbol{M}$ we may try and write the uniform measure on the space a_i. More precisely, we consider the measure

$$\mu_{\rm can}(a_i) d^2a_1 \ldots d^2a_D = 2\pi N_\mu \delta\left(Z(\beta) - \sum_i |a_i|^2 e^{-\beta E_i}\right) d^2a_1 \ldots d^2a_D. \quad (8.30)$$

Here, to make the measure well defined we had to truncate the range of energies $[E_1, E_2]$ so that the total number of eigenstates that enter the range is D. If we take this range to be large enough so that $E_2 - E_1 \gg \sqrt{\mathcal{N}}$ then, for the purposes of its action on states (8.28), the action of $\boldsymbol{M}$ can be consistently restricted to this range. Now, naively, one might believe that this leads to a contradiction with (8.25). However, we find that under (8.30) with a large range of energies the states (8.28) are themselves very atypical. Therefore the fact that the truncated version of $\boldsymbol{M}$ maps the energy range back to itself and also leaves the measure (8.30) invariant still does not lead to a contradiction with (8.25).

We now explicitly bring out the tension between measures like (8.30) which are the natural guesses for measures invariant under $\boldsymbol{M}$ and the fact that we would like the magnitudes of the a_i to be approximately constant in (8.29). We compute the reduced probability distribution, $\mu_{\rm red}$ for the coefficient a_1 by integrating out $a_2 \ldots a_D$. We write the delta function as

$$\delta\left(Z(\beta) - \sum_i |a_i|^2 e^{-\beta E_i}\right) = \lim_{\epsilon \to 0} \int \frac{dl}{2\pi} e^{il(Z(\beta) - \sum_i |a_i|^2 e^{-\beta E_i}) - \epsilon l^2},$$

where ϵ is a small regulator. We also add small regulators $\epsilon' e^{-\beta E_i} |a_i|^2$ to make the integrals over $a_2 \ldots a_D$ well defined. Then we find

$$\begin{aligned}
\mu_{\rm red}(a_1) &\equiv \int \mu_{\rm can}(a_i) d^2a_2 \ldots d^2a_D = N_\mu \int d^2a_2 \ldots d^2a_D \lim_{\epsilon,\epsilon' \to 0} \int dl e^{il(Z(\beta) - \sum_i |a_i|^2 e^{-\beta E_i}) - \epsilon l^2} e^{-\epsilon' \sum_i e^{-\beta E_i} |a_i|^2} \\
&= \left[\frac{N_\mu \pi^{D-1}}{e^{-\beta \sum_i E_i}}\right] \lim_{\epsilon,\epsilon' \to 0} \int dl \frac{e^{il(Z(\beta) - |a_1|^2 e^{-\beta E_1}) - \epsilon l^2}}{(\epsilon' + il)^{D-1}} \\
&= \left[\frac{N_\mu \pi^{D-1}}{\Gamma(D-1) e^{-\beta \sum_i E_i}}\right] \int dl dx x^{D-2} e^{-x(il+\epsilon')} e^{il(Z(\beta) - |a_1|^2 e^{-\beta E_1}) - \epsilon l^2} \\
&= \left[\frac{N_\mu \pi^{D-1}}{\Gamma(D-1) e^{-\beta \sum_i E_i}} \sqrt{\frac{\pi}{\epsilon}}\right] \int dx x^{D-2} e^{-\frac{(x + Z(\beta) - |a_1|^2 e^{-\beta E_1})^2}{4\epsilon} - x\epsilon'} \\
&= \kappa \left(1 - \frac{|a_1|^2 e^{-\beta E_1}}{Z(\beta)}\right)^{D-2}.
\end{aligned}$$

In the last step here, we have absorbed all the normalization factors into an irrelevant constant κ and taken all regulators to 0 and kept the part that is nonvanishing in this limit.

Generalizing this computation to the other coefficients, we find that the reduced probability distribution for the coefficient $|a_i|^2$ can be written as

$$\mu_{\rm red}(a_i) = \kappa\left(1 - \frac{|a_i|^2 e^{-\beta E_i}}{Z(\beta)}\right)^{D-2} \approx \kappa \exp\left[-\frac{D e^{-\beta E_i}|a_i|^2}{Z(\beta)}\right]. \tag{8.31}$$

Now, we see something interesting. If we take the range of energies $[E_1, E_2]$ that appeared in (8.28) to be much larger than $\sqrt{\mathcal{N}}$ as we would need to make $\boldsymbol{M}$ act effectively in this space then (8.31) suggests that the different a_i have very different typical magnitudes. To ensure that the typical magnitudes of the coefficients a_i are the same in (8.31), we have to take the range of energies $E_1 - E_2 \ll 1$. However, in this case the ensemble is clearly not invariant under the action of $\boldsymbol{M}$.

Physical intuition: Let us briefly summarize the physical intuition behind the analysis above. The action of $\boldsymbol{M}$ is like a unitary on the coefficients a_i. Therefore, just like unitaries in a microcanonical ensemble, $\boldsymbol{M}$ tends to move the coefficients slightly from lower to higher energies. From this point of view, in the states (8.28), as written, the high energy states are weighted with coefficients that are typically too small and the low energy states are weighted with coefficients that are typically too large. If we truncate the coefficients a_i to a small range of energies, then $\boldsymbol{M}$ simply moves us out of this range. This suggests that it may be difficult to find a measure on the Hilbert space that satisfies (8.22)–(8.23) for which $\boldsymbol{M}$ does not meet (8.25).

So, in principle it is consistent to expect that there may exist further criteria, based on the magnitudes and the phases of (8.28) which can be detected by various operators beyond the simple operators in our algebra, which will determine that in the set (8.21) at most one of the states is in equilibrium whereas the others are not. We return to this issue in future work.

D. Summary

We now summarize the results of this section.

(1) For ordinary excitations of an equilibrium state with unitary operators, we can detect them using ordinary correlators and modify the construction of our mirrors accordingly.
(2) For the van Raamsdonk-type unitaries, which act behind the horizon, we can detect them by using correlators of the Hamiltonian.
(3) Harlow attempted to define new mirrors that could evade detection by the Hamiltonian. However, we have shown here that this was predicated on an error in the computation of the Hamiltonian with the mirror operators. Harlow's operators do not have the right geometric properties to play the role of mirror operators, and do not even obey the Heisenberg equations of motion.
(4) Nevertheless, for some states with a canonical spread, we can find a group of transformations as in (8.21) so that we can map one state to another where the correlators are almost the same. There is no strict ambiguity involved here, because none of these states coincide exactly with the states obtained by acting on an equilibrium state with a mirror operator.
(5) However, while it is true that at the moment we do not know how to classify the states in the orbit (8.21), we have further shown that it is consistent with statistical mechanics expectations to classify one of these as equilibrium and the others as non-equilibrium. Although it appears that all these states are equally generic, this is specious, and such a classification would be perfectly consistent with the notion that most states are equilibrium states.

We return to this issue of the classification in further work. However, we note that this is a broader question in AdS/CFT—that of precursors. At the moment, we do not know how to write down the bulk to boundary map for all possible states but this is an issue that extends beyond our construction, and is independent of the recent discussions on the information paradox. We emphasize again that our results in this subsection show that, within the class of states we have considered—equilibrium states, near-equilibrium states excited by the ordinary and mirror operators, and small superpositions of these—there is no ambiguity in our construction.

IX. STATE-DEPENDENCE IN ENTANGLED SYSTEMS AND ER = EPR

We now describe the construction of our operators in general entangled systems. In Sec. VII F, we already examined the construction of the interior in a specific entangled state—the eternal black hole. Here we generalize the construction to more general entangled states. We show, also, that the construction of Sec. VII F follows automatically from our generalized definition here.

We first present a general construction of interior operators. This construction is a very natural generalization of the one-sided interior constructed in Sec. VII and in fact the defining equations for the mirror are unchanged. The only difference is in the construction of the little Hilbert space $\mathcal{H}_{\Psi_{\rm en}}$. This is because for entangled systems we have two sets of possible natural excitations: one, where we act with excitations in the original CFT, and the other where we act with excitations in the entangled system.

We then examine the consequences of this construction. We divide this analysis into two parts. We first consider states where the CFT is entangled with another CFT in a maximal manner so that the entanglement entropy scales with $\mathcal{N}$. Next we consider states where the CFT is entangled with a small "pointer," which could be a collection of a few qubits so that the entanglement entropy is O(1).

In both cases, we obtain interesting results. When the CFT is entangled with another CFT, our construction leads to a precise and natural formulation of the ER = EPR conjecture [13]. When light operators on the right are entangled with light operators on the left, we find that excitations on the left can affect the experience of the right-infalling observer in precisely that manner predicted by a geometric wormhole. On the other hand, in a generic state where there is no such entanglement we find that an observer on the left CFT loses his power to affect the region behind the right horizon by means of simple operations, although he could possibly do so by using some very complicated operators. This is consistent with the heuristic notion that the wormhole becomes very long for these states.

On the other hand, when the CFT is entangled with a small system no such geometric wormhole appears for any state. However, for this case, there is another crucial question, which is as follows. As we show below, the important test of whether there are any observable violations of quantum mechanics for the infalling observer arises when the observer entangles the CFT with a small system, jumps into the black hole and observes whether the state-dependence leads to any deviations from linearity. We show below that such an experiment does not lead to any observable departure from the predictions of quantum mechanics.

We wish to emphasize throughout this section that these predictions arise as a natural consequence of our construction and not because we have tailored the definition of the interior operators to entangled systems. As we mentioned above, the only change in an entangled system is that we have additional coarse or light operators to excite the system from the left and therefore we must enlarge the space $\mathcal{H}_{\Psi_{\rm en}}$.

We should mention that our emphasis and approach is complementary to the approach of directly studying density matrices that was adopted in [17].

Notation and objective: In this section, we consider entangled states,

$$|\Psi_{\rm en}\rangle = \sum_i \alpha_i |\tilde{\Psi}_i\rangle \otimes |\Psi_i\rangle. \tag{9.1}$$

Here α_i are some coefficients, $|\Psi_i\rangle$ are orthonormal states in the original CFT, and $|\tilde{\Psi}_i\rangle$ are states in a second system that may be another CFT or a collection of qubits. We refer to this system as the *left* system. The sum may be over a small number of states, or an exponentially large number.

In this section, our primary objective is to reconstruct the experience of the infalling observer from the original CFT, which we also call the *right* CFT. Our construction of the mirrors, and also the little Hilbert space is appropriate for right-relationally defined local observables. In many cases where the left system is also a CFT, we can perform an analogous construction to describe the experience of a left-infalling observer. But apart from indicating this briefly below, we do not focus on this.

A. Mirror operators for entangled systems

Summary of the construction: The construction can be summarized as follows. We call $\mathcal{A}$ the small algebra of the right CFT and $\mathcal{A}_L$ for the algebra of observables of the left system. We also define the product of the two algebras $\mathcal{A}_{\rm product} = \mathcal{A}_L \otimes \mathcal{A}$.

The little Hilbert space is defined as the span of states $\{\mathcal{A}_{\rm product}|\Psi_{\rm en}\rangle\}$. In general this is bigger than just the span of states $\{\mathcal{A}|\Psi\rangle\}$, but there are some cases (like the thermofield double state) where the two spaces are the same. In the general case, the Hilbert space $\mathcal{H}_{\Psi_{\rm en}}$ can be decomposed into the direct sum of subspaces $\mathcal{H}^j_{\Psi_{\rm en}}$, each of which is closed under the action of the right algebra $\mathcal{A}$,

$$\mathcal{H}_{\Psi_{\rm en}} = \bigoplus_j \mathcal{H}^j_{\Psi_{\rm en}}.$$

For each j we can identify a unique state $|\Psi^j_{\rm en}\rangle \in \mathcal{H}^j_{\Psi_{\rm en}}$ which is an equilibrium state with respect to the right CFT.[26] The rest of the subspace $\mathcal{H}^j_{\Psi_{\rm en}}$ can be generated by acting on this equilibrium vector with elements of the algebra $\mathcal{A}$.

Hence, within each of these subspaces we have a representation of the algebra $\mathcal{A}$ which obeys all the conditions that we encountered in the case of nonentangled systems. More precisely, no element of the algebra $\mathcal{A}$ can annihilate the state $|\Psi^j_{\rm en}\rangle$ and the entire Hilbert space $\mathcal{H}^j_{\Psi_{\rm en}}$ can be generated by acting with $\mathcal{A}$ on $|\Psi^j_{\rm en}\rangle$. The first condition follows from our assumption that right-CFT states in (9.1) are black hole states.

We can now define the mirror operators acting within this subspace using exactly the same rules as in Sec. VII. Finally, the mirror operators acting on the full little Hilbert space $\mathcal{H}_{\Psi_{\rm en}}$ are just the sums of the individual mirror operators on the subspaces $\mathcal{H}^j_{\Psi_{\rm en}}$.

We emphasize that this is the natural extension of our construction of the mirror operators for systems without entanglement. As we see, this simple definition is able to reproduce the expected physics for ER = EPR and other types of entangled states with or without wormholes. Below we describe this construction in more detail.

1. Construction of the little Hilbert space for entangled systems

We now discuss in detail how to construct the little Hilbert space about an entangled state $\mathcal{H}_{\Psi_{\rm en}}$. We first

[26] As in Sec. VII E 2 when considering superpositions, it may happen that there is no equilibrium state inside $\mathcal{H}^j_{\Psi_{\rm en}}$. In this case we need to enlarge $\mathcal{H}^j_{\Psi_{\rm en}}$ to the direct sum of little Hilbert spaces built on equilibrium states.

discuss the set of allowed excitations. We then use this to discuss the notion of equilibrium in entangled systems. Finally we put these notions together to construct $\mathcal{H}_{\Psi_{\rm en}}$.

Allowed excitations of entangled systems: There are two differences from the single-sided construction. In an entangled system, we have first the operators from the original CFT, which are part of $\mathcal{A}$. Additionally, observers should also have the ability to excite the state by acting with operators in the left system as well. In the left system, we can again build up a set of operators, which we denote by $\mathcal{A}_L$. If the left system is a holographic CFT, we should restrict the set of allowed operators in the same way that we restrict them for the original CFT. On the other hand if the left system is a collection of qubits, then there is no notion of light and heavy operators, and we can allow $\mathcal{A}_L$ to include *all* operators in the left theory. Since operators on the left commute with operators on the right the full set of allowed operators has the structure of a direct product

$$\mathcal{A}_{\rm product} = \mathcal{A}_L \otimes \mathcal{A}.$$

We denote elements of the left algebra by $\boldsymbol{A}_{L,\alpha} \in \mathcal{A}_L$, and elements of the original algebra by $\boldsymbol{A}_\alpha \in \mathcal{A}$ as usual.

We explore this in greater detail below but we caution the reader that unlike in the case of the single-sided CFT the little Hilbert space $\mathcal{H}_{\Psi_{\rm en}}$ is not isomorphic to $\mathcal{A}_{\rm product}$.

Equilibrium in entangled systems: We now turn to the notion of equilibrium in entangled systems. Since we are now allowing excitations of the state by operators in $\mathcal{A}_{\rm product}$ it is natural to modify the notion of equilibrium as well. This is a natural generalization of the definition of equilibrium in Sec. VII B for the original CFT. We define the deviation from equilibrium on the right using the same parameters as in (7.3)–(7.4),

$$\chi_p(t) = \langle\Psi_{\rm en}|e^{i\boldsymbol{H}t}\boldsymbol{A}_p e^{-i\boldsymbol{H}t}|\Psi_{\rm en}\rangle,$$
$$\nu_p = T_{\rm b}^{\frac{-1}{2}}\int_0^{T_{\rm b}} |(\chi_p(t) - \chi_p(0))|dt,$$

where $\boldsymbol{H}$ is the right Hamiltonian. In addition, we consider similar deviations from equilibrium in the left CFT.

$$\chi_{Lp}(t) = \langle\Psi_{\rm en}|e^{i\boldsymbol{H}_L t}\boldsymbol{A}_{{\rm L},p} e^{-i\boldsymbol{H}_L t}|\Psi_{\rm en}\rangle,$$
$$\nu_{Lp} = T_{\rm b}^{\frac{-1}{2}}\int_0^{T_{\rm b}} |(\chi_{Lp}(t) - \chi_{Lp}(0))|dt.$$

A necessary condition for the system to be in equilibrium is then that both left and right correlators are time-translationally invariant.

$$\nu_p = {\rm O}(e^{-\frac{S}{2}}), \quad \forall\ p,$$
$$\nu_{Lp} = {\rm O}(e^{-\frac{S}{2}}), \quad \forall\ p. \tag{9.2}$$

As above this condition is necessary but not strictly sufficient because of the class of excitations that we discussed in Sec. VIII C. We also see below that (9.2) is often superfluous and we can perform the construction of the mirrors provided that the state is in right equilibrium even if it is not in left equilibrium.

$\mathcal{H}_{\Psi_{\rm en}}$ *for entangled states*: We now turn to the construction of the little Hilbert space, which describes the space of simple excitations about the base state. The main difference compared to our discussion above is that in the presence of entanglement, it is not necessary that all operators in $\mathcal{A}_{\rm product}$ give rise to independent descendants of the state $|\Psi_{\rm en}\rangle$. In particular, it is possible that

$$(\boldsymbol{A}_{{\rm L},p} - \boldsymbol{A}_q)|\Psi_{\rm en}\rangle = 0,$$

for some correlated choices of $\boldsymbol{A}_{{\rm L},p}$ and $\boldsymbol{A}_q$. Let us consider two examples of this.

In the thermofield state $|\Psi_{\rm tfd}\rangle$, we have

$$(\mathcal{O}_{L\omega} - e^{-\frac{\beta\omega}{2}}\mathcal{O}_\omega^\dagger)|\Psi_{\rm tfd}\rangle = 0. \tag{9.3}$$

It is understood, above and in other equations below, that when we write an operator purely from the left system, it can be lifted to an operator on the product system through $\mathcal{O}_{L,\omega} \equiv \mathcal{O}_{L,\omega} \otimes \boldsymbol{1}_R$ and vice versa.

Next, consider the CFT entangled with a two qubit system. This system has four states, which we denote by $|1\rangle\ldots|4\rangle$. Now we may have a state that is not maximally entangled,

$$|\Psi_{\rm en}\rangle = \frac{1}{\sqrt{3}}(|\Psi_1\rangle \otimes |1\rangle + |\Psi_2\rangle \otimes |2\rangle + |\Psi_3\rangle \otimes |3\rangle),$$

where $|\Psi_i\rangle$ are some orthogonal states in the original CFT. Denoting the projector onto state $|4\rangle$ by $P_4 = |4\rangle\langle 4|$, we see clearly that

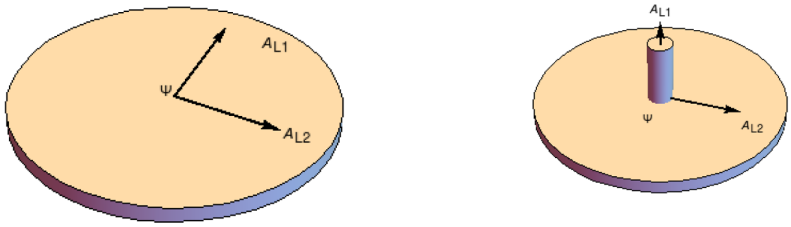
$$P_4|\Psi_{\rm en}\rangle = 0. \tag{9.4}$$

Note that both these kinds of states, where we obtain null relations, are very special. States where relations of the form (9.3) hold are special because the entanglement is between simple operators on both sides. As we see below, generic states do not have such relations. Similarly, when the left system is small, relations of the form (9.4) also occur only when the entanglement is nonmaximal. Nevertheless, our construction is able to account for these null relations correctly.

We now define $\mathcal{H}_{\Psi_{\rm en}}$ as follows. Starting with the state $|\Psi_{\rm en}\rangle$, we act with all elements of $\mathcal{A}$ to obtain the space

$$\mathcal{H}^0_{\Psi_{\rm en}} = \text{span of}\{\boldsymbol{A}_1|\Psi_{\rm en}\rangle, \ldots \boldsymbol{A}_{\mathcal{D}}|\Psi_{\rm en}\rangle\}, \tag{9.5}$$

where we remind the reader that the elements of $\mathcal{A}$ displayed above form a complete basis for this linear set. As usual we assume that there are no null vectors in the set displayed in (9.5). We define $\boldsymbol{P}^{\boldsymbol{0}}_{\rm en}$ to be the projector onto this subspace. This means that



(a) $\mathcal{H}_{\Psi_{en}}$ where the action of the "left algebra" is entirely contained within the space obtained by acting with the right algebra.

(b) $\mathcal{H}_{\Psi_{en}}$ in cases with less entanglement. Now the action of "left operators" opens up new directions.

FIG. 12. The structure of the wormhole is directly linked to the structure of $\mathcal{H}_{\Psi_{en}}$. In the case on the left above, where $\mathcal{H}_{\Psi_{en}}$ coincides with $\mathcal{H}^0_{\Psi_{en}}$, we obtain a geometric wormhole. The case on the right can be understood as an elongated wormhole. In the extreme case where $\mathcal{H}_{\Psi_{en}}$ becomes a direct product space, the geometric wormhole disappears.

$$|v\rangle \in \mathcal{H}^0_{\Psi_{en}} \Rightarrow \boldsymbol{P}^{\boldsymbol{0}}_{en}|v\rangle = |v\rangle,$$
$$\langle v|\boldsymbol{A}_p|\Psi_{en}\rangle = 0, \quad \forall\ p \Rightarrow \boldsymbol{P}^{\boldsymbol{0}}_{en}|v\rangle = 0.$$

Next we pick a Hermitian element, $\boldsymbol{A}_{L,1}$ of $\mathcal{A}_L$, and construct

$$|\Psi^1_{en}\rangle = (1 - \boldsymbol{P}^{\boldsymbol{0}}_{en})\boldsymbol{A}_{L,1}|\Psi_{en}\rangle. \tag{9.6}$$

We pick $\boldsymbol{A}_{L,1}$ so that $|\Psi^1_{en}\rangle$ is nonvanishing and in *right equilibrium*. Note that it is not necessary for $|\Psi^1_{en}\rangle$ to be in left equilibrium. (The reason for the restriction that $\boldsymbol{A}_{L,1}$ be Hermitian is explained below.) We now construct the space

$$\mathcal{H}^1_{\Psi_{en}} = \text{span of}\{\boldsymbol{A}_1|\Psi^1_{en}\rangle, \ldots \boldsymbol{A}_{\mathcal{D}}|\Psi^1_{en}\rangle\}. \tag{9.7}$$

Then we define $\boldsymbol{P}^{\boldsymbol{1}}_{en}$ to be the projector on $\mathcal{H}^1_{\Psi_{en}}$. Similarly, we look for $\boldsymbol{A}_{L,2} \in \mathcal{A}_L$ so that

$$|\Psi^2_{en}\rangle = (1 - \boldsymbol{P}^{\boldsymbol{0}}_{en})(1 - \boldsymbol{P}^{\boldsymbol{1}}_{en})\boldsymbol{A}_{L,2}|\Psi_{en}\rangle$$

is nonvanishing and in right equilibrium. We then construct $\mathcal{H}^2_{\Psi_{en}}$ analogously to (9.5) and (9.7) and continue recursively in this manner until it is no longer possible to find any elements of $\mathcal{A}_L$ which can produce descendants of $|\Psi_{en}\rangle$ that are orthogonal to all the previous subspaces.

To summarize this construction, we find elements $\boldsymbol{A}_{L,1}\ldots\boldsymbol{A}_{L,\mathcal{D}_{max}}$ (where $\mathcal{D}_{max}$ may be smaller than the dimension of the left algebra) with the property that

$$\boldsymbol{A}_{L,1}|\Psi_{en}\rangle\ldots\boldsymbol{A}_{L,\mathcal{D}_{max}}|\Psi_{en}\rangle$$

are all in right equilibrium and have the property that

$$\langle\Psi_{en}|\boldsymbol{A}_p\boldsymbol{A}_{L,j}|\Psi_{en}\rangle = 0, \quad \forall\ p, j.$$

On each of these we construct the space $\mathcal{H}^m_{\Psi_{en}}$ as shown in (9.5) and (9.7). The full space $\mathcal{H}_{\Psi_{en}}$ is then defined by

$$\mathcal{H}_{\Psi_{en}} = \bigoplus_j \mathcal{H}^j_{\Psi_{en}}.$$

It is worth discussing the structure of the space $\mathcal{H}_{\Psi_{en}}$ that results from the construction above, and the examples that we consider below will elucidate this. In the thermofield state, an action by a simple operator in the left CFT corresponds to the action of a simple operator on the right CFT. Therefore in this case $\mathcal{H}_{\Psi_{en}}$ coincides with $\mathcal{H}^0_{\Psi_{en}}$. On the other hand, in a generic entangled state of two CFTs, there is no relation between the action of simple operators on the left and the right, and therefore $\mathcal{H}_{\Psi_{en}}$ is isomorphic to $\mathcal{A} \otimes \mathcal{A}_{\text{product}}$. In intermediate cases where there is some entanglement, but not maximal, we obtain an $\mathcal{H}_{\Psi_{en}}$ that is intermediate between these two cases: its dimension is larger than $\mathcal{H}^0_{\Psi_{en}}$ but not maximal. We describe this in detail in several cases below.

The structure of $\mathcal{H}_{\Psi_{en}}$ is directly related to whether we obtain a wormhole on this. This is shown schematically in Fig. 12 and explained further below.

Definition of the mirror operators: The mirror operators are now defined via precisely the same linear equations as Sec. VII C. Note that each vector in $\mathcal{H}_{\Psi_{en}}$ can be written as a linear combinations of vectors of the form $\boldsymbol{A}_p|\Psi^j_{en}\rangle$ for some choice of p and j. We define

$$\tilde{\boldsymbol{\mathcal{O}}}_{\omega,m}\boldsymbol{A}_p|\Psi^j_{en}\rangle = \boldsymbol{A}_p e^{-\frac{\beta\omega}{2}}(\boldsymbol{\mathcal{O}}_{\omega,m})^\dagger|\Psi^j_{en}\rangle,$$
$$[\tilde{\boldsymbol{\mathcal{O}}}_{\omega,m}, \boldsymbol{H}]\boldsymbol{A}_p|\Psi^j_{en}\rangle = -\omega\tilde{\boldsymbol{\mathcal{O}}}_{\omega,m}\boldsymbol{A}_p|\Psi^j_{en}\rangle. \tag{9.8}$$

As usual, these equations have a solution because we have $\boldsymbol{A}_p|\Psi^j_{en}\rangle \neq 0,\ \forall\ p, j$. As the reader will note this is a direct extension of our definition of the mirrors for the original CFT. We now show how this simple extension has remarkable properties and allows us to derive a precise version of the ER = EPR conjecture and also show that the infalling observer will not observe any violations of quantum mechanics.

B. The wormhole in the thermofield double state

We now show how the construction above leads to a wormhole in the thermofield double state, where we take $|\Psi_{\rm en}\rangle = |\Psi_{\rm tfd}\rangle$. First, let us examine the construction of $\mathcal{H}_{\Psi_{\rm en}}$. In the thermofield state we have the following relations:

$$\begin{aligned}\mathcal{O}_{L\omega,m}|\Psi_{\rm tfd}\rangle &= e^{-\frac{\beta\omega}{2}}\mathcal{O}^\dagger_{\omega,m}|\Psi_{\rm tfd}\rangle,\\ \mathcal{O}^\dagger_{L\omega,m}|\Psi_{\rm tfd}\rangle &= e^{\frac{\beta\omega}{2}}\mathcal{O}_{\omega,m}|\Psi_{\rm tfd}\rangle.\end{aligned} \tag{9.9}$$

Now consider an arbitrary polynomial in the $\mathcal{O}_{L\omega,m}$, which we denote by $\boldsymbol{A}_{{\rm L},\alpha}$. In the thermofield state we have the relation

$$\boldsymbol{A}_{{\rm L},\alpha}|\Psi_{\rm tfd}\rangle = e^{-\frac{\beta\boldsymbol{H}}{2}}\boldsymbol{A}^\dagger_\alpha e^{\frac{\beta\boldsymbol{H}}{2}}|\Psi_{\rm tfd}\rangle,$$

where, on the right of the equation above, we have an operator acting purely in the right CFT. If $\boldsymbol{A}_{{\rm L},\alpha}\in\mathcal{A}_L$ then, barring edge effects, we have $e^{-\frac{\beta\boldsymbol{H}}{2}}\boldsymbol{A}^\dagger_\alpha e^{\frac{\beta\alpha}{2}}\in\mathcal{A}$. Therefore, in this case we start by constructing

$$\mathcal{H}^0_{\Psi_{\rm tfd}} = \mathcal{A}|\Psi_{\rm tfd}\rangle,$$

and then we do not get any new states by acting with $\mathcal{A}_L$. As a result, the full little Hilbert space is simply

$$\mathcal{H}_{\Psi_{\rm tfd}} = \mathcal{H}^0_{\Psi_{\rm tfd}}.$$

Then the construction of the mirror operators results in the same answer as the construction in Sec. VII F but we repeat it here from the general perspective of mirrors in entangled systems that we have presented above. The action of the mirror operators is specified by the linear equations (9.8). Since in this case the structure of $\mathcal{H}_{\Psi_{\rm tfd}}$ is so simple, these equations reduce to

$$\begin{aligned}\tilde{\mathcal{O}}_{\omega,m}\boldsymbol{A}_\alpha|\Psi_{\rm tfd}\rangle &= \boldsymbol{A}_\alpha e^{-\frac{\beta\omega}{2}}\mathcal{O}^\dagger_{\omega,m}|\Psi_{\rm tfd}\rangle,\\ [\tilde{\mathcal{O}}_{\omega,m},\boldsymbol{H}]\boldsymbol{A}_\alpha|\Psi_{\rm tfd}\rangle &= -\omega\boldsymbol{A}_\alpha e^{-\frac{\beta\omega}{2}}\mathcal{O}^\dagger_{\omega,m}|\Psi_{\rm tfd}\rangle.\end{aligned} \tag{9.10}$$

Now the first point we note is that $\tilde{\mathcal{O}}_{\omega,m}$ does not *commute* with elements of $\mathcal{A}_L$, and moreover that this nonzero commutator is very special. We can check this explicitly by considering the commutator of $[\tilde{\mathcal{O}}_{\omega,m},\mathcal{O}^\dagger_{L\omega',m'}]$. We have

$$\begin{aligned}\tilde{\mathcal{O}}_{\omega,m}\mathcal{O}^\dagger_{L\omega',m'}|\Psi_{\rm tfd}\rangle &= e^{\frac{\beta\omega'}{2}}\tilde{\mathcal{O}}_{\omega,m}\mathcal{O}_{\omega',m'}|\Psi_{\rm tfd}\rangle\\ &= e^{\frac{\beta(\omega'-\omega)}{2}}\mathcal{O}_{\omega',m'}\mathcal{O}^\dagger_{\omega,m}|\Psi_{\rm tfd}\rangle,\end{aligned}$$

where in the first equality we used (9.9). And also

$$\begin{aligned}\mathcal{O}^\dagger_{L\omega',m'}\tilde{\mathcal{O}}_{\omega,m}|\Psi_{\rm tfd}\rangle &= e^{-\frac{\beta\omega}{2}}\mathcal{O}^\dagger_{L\omega',m'}\mathcal{O}^\dagger_{\omega,m}|\Psi_{\rm tfd}\rangle\\ &= e^{-\frac{\beta\omega}{2}}\mathcal{O}^\dagger_{\omega,m}\mathcal{O}^\dagger_{L\omega',m'}|\Psi_{\rm tfd}\rangle\\ &= e^{\frac{\beta(\omega'-\omega)}{2}}\mathcal{O}^\dagger_{\omega,m}\mathcal{O}_{\omega',m'}|\Psi_{\rm tfd}\rangle.\end{aligned}$$

This leads to an O(1) effective commutator,

$$[\tilde{\mathcal{O}}_{\omega,m},\mathcal{O}^\dagger_{L\omega',m'}]|\Psi_{\rm tfd}\rangle = C_\beta(\omega,m)\delta_{\omega\omega'}\delta_{mm'}|\Psi_{\rm tfd}\rangle. \tag{9.11}$$

These are very special commutators, and suggest that within correlators involving only elements of $\mathcal{A}_{\rm gff}$, it is possible to replace $\tilde{\mathcal{O}}_{\omega,m}$ with $\mathcal{O}_{L\omega,m}$. However, as we have emphasized one cannot equate these operators. In particular, to compute the commutator of the mirrors with the left Hamiltonian we consider

$$\begin{aligned}\tilde{\mathcal{O}}_{\omega,m}\boldsymbol{H}_L|\Psi_{\rm tfd}\rangle = \tilde{\mathcal{O}}_{\omega,m}\boldsymbol{H}|\Psi_{\rm tfd}\rangle &= e^{-\frac{\beta\omega}{2}}\mathcal{O}^\dagger_{\omega,m}\boldsymbol{H}|\Psi_{\rm tfd}\rangle\\ &= e^{-\frac{\beta\omega}{2}}\mathcal{O}^\dagger_{\omega,m}\boldsymbol{H}_L|\Psi_{\rm tfd}\rangle\\ &= \boldsymbol{H}_L e^{-\frac{\beta\omega}{2}}\mathcal{O}^\dagger_{\omega,m}|\Psi_{\rm tfd}\rangle = \boldsymbol{H}_L\tilde{\mathcal{O}}_{\omega,m}|\Psi_{\rm tfd}\rangle.\end{aligned}$$

In this chain of equalities we have first used the isometry of the thermofield state, then used the definition (9.10) and then manipulated this expression by using the isometry again and the fact that $\boldsymbol{H}_L$ commutes with right operators. So we find that within simple correlators

$$[\tilde{\mathcal{O}}_{\omega,m},\boldsymbol{H}_L]|\Psi_{\rm tfd}\rangle \doteq 0.$$

Therefore the mirror operators have a vanishing commutator with the left Hamiltonian. Note that this follows as a *consequence* of our defining relations and is not something that we have to put in by hand.

For the sake of completeness, we can also evaluate the two-point function

$$\begin{aligned}\langle\Psi_{\rm tfd}|\tilde{\mathcal{O}}_{\omega,m}\mathcal{O}^\dagger_{L\omega,m}|\Psi_{\rm tfd}\rangle &= e^{\frac{\beta\omega}{2}}\langle\Psi_{\rm tfd}|\tilde{\mathcal{O}}_{\omega,m}\mathcal{O}_{\omega,m}|\Psi_{\rm tfd}\rangle\\ &= G_\beta(\omega,m).\end{aligned} \tag{9.12}$$

We can proceed to evaluate other correlators along the lines of (9.11)–(9.12). If we now try and reproduce these correlators from a geometry then the geometric picture that arises from this is that of the standard thermofield wormhole. See Fig. 13. Now we show how, in a generic entangled state of the two CFTs, a very different geometric picture emerges.

C. The generic entangled state of two CFTs

We now show how our construction works in the generic entangled state of two CFTs. Consider scrambling the thermofield double state with a left unitary. So we now consider

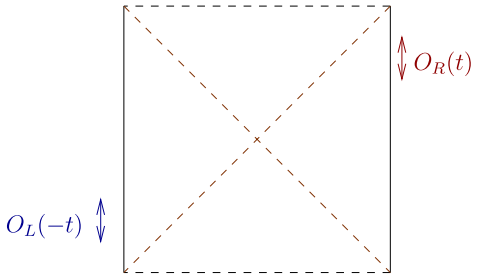


FIG. 13. The standard wormhole described in Sec. IX B: operators on the right $\mathcal{O}_R(t)$ are entangled with left operator $\mathcal{O}_L(-t)$.

$$|\Psi_{\rm gen}\rangle = \boldsymbol{U}_{L,g}|\Psi_{\rm tfd}\rangle, \tag{9.13}$$

where the unitary is *not* an exponentiated element of the algebra of simple operator: $\boldsymbol{U}_{L,g} \neq e^{i\boldsymbol{A}_{{\rm L},\alpha}}$, but rather some generic unitary that changes the structure of entanglement of the two sides. As a result, as shown in [4], simple operators on the left and right are uncorrelated.

$$\langle\Psi_{\rm tfd}|\boldsymbol{U}^\dagger_{L,g}\boldsymbol{A}_{{\rm L},\alpha}\boldsymbol{A}_\beta\boldsymbol{U}_{L,g}|\Psi_{\rm tfd}\rangle = {\rm O}(e^{-\frac{S}{2}}), \quad \forall\ \alpha,\beta. \tag{9.14}$$

The construction of $\mathcal{H}_{\Psi_{\rm gen}}$ proceeds according to the algorithm described in the beginning of this section. Notice that there is a qualitative difference from the thermofield double state, because we no longer have relations of the form (9.9). The relation (9.14) implies that for an arbitrary element $\boldsymbol{A}_{\rm L,1} \in \mathcal{A}_L$, the left descendant constructed via (9.6) is non-null and in right equilibrium. Hence the little Hilbert space $\mathcal{H}_{\Psi_{\rm gen}}$ will have the direct sum decomposition as explained earlier. We select a set of operators $\boldsymbol{A}_{\rm L,1}\ldots\boldsymbol{A}_{{\rm L},\mathcal{D}_L}$ which form a basis of $\mathcal{A}_L$ and generate the equilibrium vector in each of these subspaces. Finally we find

$$\mathcal{H}_{\Psi_{\rm gen}} = {\rm span\ of}\{\boldsymbol{A}_\beta\boldsymbol{A}_{{\rm L},\alpha}|\Psi_{\rm gen}\rangle, \\ \beta = 1\ldots\mathcal{D}; \alpha = 1\ldots\mathcal{D}_L\}.$$

Now, the definition of the mirror operators above reads

$$\tilde{\mathcal{O}}_{\omega,m}\boldsymbol{A}_\beta\boldsymbol{A}_{{\rm L},\alpha}|\Psi_{\rm gen}\rangle = \boldsymbol{A}_\beta e^{-\frac{\beta\omega}{2}}\mathcal{O}^\dagger_{\omega,m}\boldsymbol{A}_{{\rm L},\alpha}|\Psi_{\rm gen}\rangle. \tag{9.15}$$

But since operators in $\mathcal{A}$ and $\mathcal{A}_L$ commute this becomes

$$\tilde{\mathcal{O}}_{\omega,m}\boldsymbol{A}_\beta\boldsymbol{A}_{{\rm L},\alpha}|\Psi_{\rm gen}\rangle = e^{-\frac{\beta\omega}{2}}\boldsymbol{A}_\beta\boldsymbol{A}_{{\rm L},\alpha}\mathcal{O}^\dagger_{\omega,m}|\Psi_{\rm gen}\rangle.$$

Therefore for the generic entangled state $|\Psi_{\rm gen}\rangle$, we have

$$[\tilde{\mathcal{O}}_{\omega,m},\boldsymbol{A}_{{\rm L},\alpha}]|\Psi_{\rm gen}\rangle = 0, \quad \text{generic state.} \tag{9.16}$$

We can also compute the two-point function

$$\langle\Psi_{\rm gen}|\tilde{\mathcal{O}}_{\omega,m}\mathcal{O}^\dagger_{L\omega,m}|\Psi_{\rm gen}\rangle = e^{\frac{-\beta\omega}{2}}\langle\Psi_{\rm gen}|\mathcal{O}^\dagger_{L\omega,m}\mathcal{O}^\dagger_{\omega,m}|\Psi_{\rm gen}\rangle \\ = {\rm O}(e^{\frac{-S}{2}}). \tag{9.17}$$

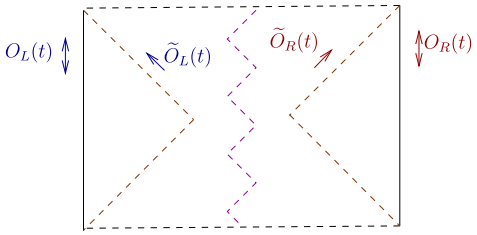


FIG. 14. The dual to the generic entangled state described in Sec. IX C. Simple operators on the right $\mathcal{O}_R(t)$ and left are not correlated. This is indicated by the jagged broken line in the middle and there is no geometric wormhole. But both sides see a smooth horizon with the emergence of new mirror operators behind the horizon.

Other two-point functions of simple operators vanish in the same manner. Therefore the mirrors not only effectively commute, they are also uncorrelated with the simple left operators.

Note that both (9.16)–(9.17)—just like (9.11)—came automatically from our definition of the mirror operators for entangled systems and the different structure of $\mathcal{H}_{\Psi_{\rm gen}}$ in these cases, without having to put anything in by hand.

Now, we may try and write down a geometry that reproduces (9.17) and (9.16). We remind the reader that correlators between the mirror operators and ordinary operators are unchanged showing that the right-infalling observer still perceives a smooth horizon. However, the vanishing commutator (9.16) shows that in the generic state it is not possible to affect the experience of the right-infalling observer by simple operators on the left. Hence the geometric wormhole has disappeared. Instead, geometrically we obtain the Penrose diagram of Fig. 14. This Penrose diagram was also conjectured in [61].

1. Mirrors as scrambled left operators in the generic state

We conclude with a further observation on the mirror operators in the generic state $|\Psi_{\rm gen}\rangle$. The relation (9.16) is somewhat deceptive. Our construction automatically leads to the conclusion that the commutator of the mirror operators for the right-infalling observer and simple left operators, where simple is defined through membership in $\mathcal{A}_L$, vanishes when inserted in low point correlation functions. However, another interesting consequence is that when we have a high degree of entanglement of the CFT with another system, then generically the mirror operators act on the left system as well. This follows as an inevitable consequence of their defining equations. It is easy to prove this as follows.

Let us write the generic entangled state in a Schmidt basis so that

$$|\Psi_{\rm gen}\rangle = \sum_i \kappa_i|\tilde{v}_i\rangle \otimes |v_i\rangle, \tag{9.18}$$

where the κ_i are arbitrary coefficients and we have diagonalized the entanglement so that $|v_i\rangle$ are some orthonormal states in the right CFT and $|\tilde{v}_i\rangle$ are some states in the left CFT. Now consider just one of the defining equations for $\tilde{\mathcal{O}}_{\omega,m}$,

$$\tilde{\mathcal{O}}_{\omega,m}\boldsymbol{A}_\alpha|\Psi_{\rm gen}\rangle = \boldsymbol{A}_\alpha e^{-\frac{\beta\omega}{2}}\mathcal{O}^\dagger_{\omega,m}|\Psi_{\rm gen}\rangle, \tag{9.19}$$

and look for a solution to (9.19) with the $\tilde{\mathcal{O}}_{\omega,m}$ acting entirely within the Hilbert space of the right CFT. We emphasize that (9.19) is just a special case of (9.15) with the element of the left algebra that appears there set to the identity. Let us denote this putative solution by $\boldsymbol{X} = \tilde{\mathcal{O}}_{\omega,m}$.

We see that this demand that $\boldsymbol{X}$ is an operator in the right CFT means that for each α, the single equation (9.19) leads to a *system* of linear equations given by

$$\boldsymbol{X}\boldsymbol{A}_\alpha|v_i\rangle = \boldsymbol{A}_\alpha e^{\frac{-\beta\omega}{2}}\mathcal{O}^\dagger_{\omega,m}|v_i\rangle, \quad \forall\ \alpha, i. \tag{9.20}$$

However, if the set i in (9.18) runs over a large enough range, then in general (9.20) has no solutions. For example, consider the situation where the states $|v_i\rangle$ provide a basis of the Hilbert space. Then, with $\boldsymbol{A}_\alpha \in \mathcal{A}_{\rm gff}$, the states

$$|w_{\alpha,i}\rangle = \boldsymbol{A}_\alpha|v_i\rangle$$

provide an *overcomplete* basis for the space if we span over all i and all α. Therefore in (9.20) we are trying to specify the action of the putative purely right mirror operator on an overcomplete basis and this is not possible in general.

For example, we can find coefficients $z_{\alpha i}$ so that

$$\sum_{\alpha,i} z_{\alpha i}\boldsymbol{A}_\alpha|v_i\rangle = 0,$$

and in general it will not be the case that (9.20) map this vector to 0. In particular on this vector we would find

$$0 = \boldsymbol{X}\sum_{\alpha,i} z_{\alpha i}\boldsymbol{A}_\alpha|v_i\rangle = \sum_{\alpha,i} z_{\alpha i}e^{-\frac{\beta\omega}{2}}\boldsymbol{A}_\alpha\mathcal{O}^\dagger_{\omega,m}|v_i\rangle \neq 0?$$

Here we have used the fact that generically the right-hand side of the relation above will not vanish with the same coefficients $z_{\alpha,i}$.

So we have shown that in the situation with a high entanglement entropy the $\tilde{\mathcal{O}}_{\omega,m}$ operators must act on the left as well and the operator $\boldsymbol{X}$ that acts only in the right CFT does not exist.

We conclude with some speculative comments on the possible physical implications of this fact. The authors of [13] suggested that the generic state $|\Psi_{\rm gen}\rangle$ may nevertheless be understood through a very long wormhole. Now note that our discussion of the generic commutator in Sec. V D suggests that if we take a *generic* operator in the left CFT, $\boldsymbol{Y}$, then we would find that

$$\langle\Psi_{\rm gen}\|[\boldsymbol{Y}, \tilde{\mathcal{O}}_{\omega,m}]|^2|\Psi_{\rm gen}\rangle = {\rm O}(1). \tag{9.21}$$

We emphasize that $\boldsymbol{Y}$ is not one of the simple operators that are part of $\mathcal{A}_L$ which commute with the mirrors within low point correlators. Now (9.21) suggests that with a suitably complicated operation the left observer can affect the experience of the right-infalling observer. This may be taken as some evidence of the existence of a long wormhole although it would be nice to make this more precise.

D. A superposition of the thermofield and a generic state

As a further example, we now show how our construction works in the superposition of the thermofield and a generic state. We consider

$$|\Psi_s\rangle = \kappa(|\Psi_{\rm tfd}\rangle + |\Psi_{\rm gen}\rangle). \tag{9.22}$$

For the generic left unitary of the sort discussed in (9.13), we have $\kappa = \frac{1}{\sqrt{2}} + {\rm O}(e^{-S})$.

We start with

$$\mathcal{H}^0_{\Psi_s} = \mathcal{A}|\Psi_s\rangle.$$

On the other hand, on acting with an element of $\mathcal{A}_L$ we find that

$$\begin{aligned}|\Psi^1_s\rangle &= (1-\boldsymbol{P}^0_s)\boldsymbol{A}_{{\rm L},1}|\Psi_s\rangle \\ &= \kappa(1-\boldsymbol{P}^0_s)(e^{-\frac{\beta H}{2}}\boldsymbol{A}^\dagger_1 e^{\frac{\beta H}{2}}|\Psi_{\rm tfd}\rangle + \boldsymbol{A}_{{\rm L},1}|\Psi_{\rm gen}\rangle) \\ &= \kappa\boldsymbol{A}_{{\rm L},1}|\Psi_{\rm gen}\rangle - \frac{1}{2}\kappa\langle\boldsymbol{A}_{{\rm L},1}\rangle(|\Psi_{\rm gen}\rangle + |\Psi_{\rm tfd}\rangle) \\ &\quad + \frac{\kappa}{2}e^{-\frac{\beta H}{2}}\boldsymbol{A}^\dagger_1 e^{\frac{\beta H}{2}}(|\Psi_{\rm tfd}\rangle - |\Psi_{\rm gen}\rangle).\end{aligned} \tag{9.23}$$

Here $\langle\boldsymbol{A}_{{\rm L},1}\rangle \equiv \langle\Psi_{\rm gen}|\boldsymbol{A}_{{\rm L},1}|\Psi_{\rm gen}\rangle$. In deriving this result, we have used two intermediate results.

$$\begin{aligned}\boldsymbol{P}^0_s(\boldsymbol{A}_{{\rm L},1} - \langle\boldsymbol{A}_{{\rm L},1}\rangle)|\Psi_{\rm gen}\rangle &= 0, \\ \boldsymbol{P}^0_s\boldsymbol{A}_m|\Psi_{\rm gen}\rangle &= \boldsymbol{P}^0_s\boldsymbol{A}_m|\Psi_{\rm tfd}\rangle \\ &= \frac{1}{2}(\boldsymbol{A}_m|\Psi_{\rm gen}\rangle + \boldsymbol{A}_m|\Psi_{\rm tfd}\rangle),\end{aligned}$$

where $\boldsymbol{A}_m$ is any element of $\mathcal{A}$.

In the final expression in (9.23) we have, once again, a superposition of an equilibrium and a near-equilibrium state from the point of view of observables in $\mathcal{A}$. This is a special case of the superposition of near-equilibrium states that was considered in Sec. VII. In such states, as explained there, we must enlarge the little Hilbert space slightly and upon doing that we find

$$\mathcal{H}_{\Psi_s} = \mathcal{H}_{\Psi_{\rm tfd}} \oplus \mathcal{H}_{\Psi_{\rm gen}}.$$

The action of the mirror operators can be deduced in a straightforward way from the definition provided in (9.8).

$$\tilde{\mathcal{O}}_{\omega,m}\boldsymbol{A}_{{\rm L},\alpha}\boldsymbol{A}_\beta|\Psi_s\rangle = \kappa \boldsymbol{A}_\beta e^{-\frac{\beta \boldsymbol{H}}{2}}\boldsymbol{A}_\alpha^\dagger e^{\frac{\beta \boldsymbol{H}}{2}} e^{-\frac{\beta\omega}{2}}\mathcal{O}^\dagger_{\omega,m}|\Psi_{\rm tfd}\rangle + \kappa \boldsymbol{A}_\beta \boldsymbol{A}_{{\rm L},\alpha} e^{-\frac{\beta\omega}{2}}\tilde{\mathcal{O}}^\dagger_{\omega,m}|\Psi_{\rm gen}\rangle.$$

Consequently correlators involving mirrors and ordinary operators separate into

$$\langle\Psi_s|\tilde{\boldsymbol{A}}_{\alpha_3}\boldsymbol{A}_{{\rm L},\alpha_2}\boldsymbol{A}_{\alpha_1}|\Psi_s\rangle = |\kappa|^2(\langle\Psi_{\rm tfd}|\tilde{\boldsymbol{A}}_{\alpha_3}\boldsymbol{A}_{{\rm L},\alpha_2}\boldsymbol{A}_{\alpha_1}|\Psi_{\rm tfd}\rangle + \langle\Psi_{\rm gen}|\tilde{\boldsymbol{A}}_{\alpha_3}\boldsymbol{A}_{{\rm L},\alpha_2}\boldsymbol{A}_{\alpha_1}|\Psi_{\rm gen}\rangle).$$

Therefore the superposition of states (9.22) acts like a classical mixture of a thermofield and a state with no wormhole. This is precisely what is expected. Note that standard Penrose diagrams cannot capture this superposition of two geometries, although the correlators are very simply related to the correlators in the two individual geometries.

E. The microcanonical double state and a low-pass wormhole

We now consider a modification of the thermofield state: a microcanonical double state. We show that in the appropriate regime this leads to a new kind of wormhole with interesting properties.

Consider a range of energy $E \pm \Delta$ that contains $\mathcal{D}_{E,\Delta}$ states. Here $\Delta = {\rm O}(1)$. It is also useful to consider energies that are high enough so that the associated temperature satisfies $\beta\Delta \ll 1$. These are all hierarchies between O(1) quantities and neither β nor Δ scale with $\mathcal{N}$. Now consider

$$|\Psi_{\rm md}\rangle = \frac{1}{\sqrt{\mathcal{D}_E}}\sum_{E_i=E-\Delta}^{E_i=E+\Delta}|E_i,E_i\rangle. \tag{9.24}$$

This state was also considered in [12] (see page 15), but we reach a conclusion that is different from the conclusion reached there. In particular, the state (9.24) does have a smooth interior and, contrary to the suggestion made in [12], our construction generates it correctly. The error made in [12] follows from the error alluded to in Sec. VIII B: an incorrect expectation that the mirror operators must correspond to simple operators in the left CFT.

Consider a frequency $\omega_l \ll \Delta$. The subscript indicates that this is a low frequency. For correlators involving such modes, the fact that the entanglement has been truncated is invisible. Let us denote the matrix elements of this operator in the energy eigenbasis by c_{ji} as in (6.11) so that we have

$$\sum_{E_i=E-\Delta}^{E_i=E+\Delta}\mathcal{O}_{\omega_l,m}|E_i,E_i\rangle = \sum_{E_i=E-\Delta}^{E_i=E+\Delta}\sum_{E_j} c_{ji}|E_i,E_j\rangle.$$

Note that, as we explained around (6.11), we can choose these matrix elements c_{ji} to be real because of the T-invariance of the modes of local operators. While the sum over j above technically runs over all energies, since we know that the matrix elements c_{ji} should be peaked around $E_i - E_j = \omega_l$, we can write

$$\mathcal{O}_{\omega_l}|\Psi_{\rm md}\rangle = \frac{1}{\sqrt{\mathcal{D}_E}}\sum_{E_i=E-\Delta}^{E_i=E+\Delta}\left[\sum_{E_j=E-\Delta-\omega_l}^{E_j=E+\Delta-\omega_l} c_{ji}|E_i,E_j\rangle\right].$$

Now, notice that we also have

$$\sum_{E_i=E-\Delta}^{E_i=E+\Delta}\mathcal{O}^\dagger_{L\omega_l,m}|E_i,E_i\rangle = \sum_{E_i=E-\Delta}^{E_i=E+\Delta}\left[\sum_{E_j=E-\Delta+\omega_l}^{E_j=E+\Delta+\omega_l} c_{ij}|E_j,E_i\rangle\right] = \sum_{E_i=E-\Delta+\omega_l}^{E_i=E+\Delta+\omega_l}\sum_{E_j=E-\Delta}^{E_j=E+\Delta} c_{ji}|E_i,E_j\rangle.$$

In the last step, we have interchanged i and j above to bring it into a form where we can compare it with the action of the right operator. However, the ranges of the sums over i, j are different. In the case where $\omega_l \ll \Delta$ and $\beta\omega_l \ll 1$ we can approximately neglect this to obtain

$$\mathcal{O}_{\omega_l,m}|\Psi_{\rm md}\rangle = \mathcal{O}^\dagger_{L\omega_l,m}|\Psi_{\rm md}\rangle + {\rm O}\left(\frac{\omega_l}{\Delta}\right) + {\rm O}(\beta\omega_l), \qquad \omega_l \ll \Delta. \tag{9.25}$$

On the other hand, for large $\omega_h \gg \Delta$ we see that

$$\langle\Psi_{\rm md}|\mathcal{O}^\dagger_{L\omega_h}\mathcal{O}_{\omega_h}|\Psi_{\rm md}\rangle \ll 1, \qquad \omega_h \gg \Delta. \tag{9.26}$$

Note that the result (9.26) holds even if $\beta\omega_h \ll 1$.

We can now perform the construction above to define the right-relational mirrors on this state. The relations (9.26) and (9.25) then tell us that inside correlation functions evaluated on (9.24) (except those involving the Hamiltonian, where $\frac{1}{\mathcal{N}}$ corrections are important) we can approximately perform the replacement for low frequencies,

$$\tilde{\mathcal{O}}_{\omega_l,m} \to \mathcal{O}_{L\omega_l,m}, \qquad \omega_l \ll \Delta.$$

However, no such replacement is possible for high frequency modes $\tilde{\mathcal{O}}_{\omega_h,m}$, which cannot be related to the action of simple left operators. These are independent operators that can be constructed using the algorithm that we have outlined. Using this we can compute correlators involving both ordinary operators on the left and the right, and the mirror operators precisely.

It would be interesting to develop a more precise picture of the geometric dual to this state. However, some qualitative properties are clear. The state (9.24) is a low-pass

wormhole—where low frequency modes on the left and right are entangled, but the mirrors for high frequency modes on both sides are independent operators. In this geometry both the left- and the right-infalling observer see smooth horizons. These observers can "communicate" using low frequencies but not high frequencies.

It may also be possible to think of these wormholes as "elongated wormholes." It is interesting to notice that the geometries described in [62], which were also considered in [13] have somewhat similar properties. However, these geometries involve infalling matter and cannot be a precise dual to $|\Psi_{\rm md}\rangle$, since the state $|\Psi_{\rm md}\rangle$ is invariant under $e^{i(\boldsymbol{H}_L-\boldsymbol{H})T}|\Psi_{\rm md}\rangle = |\Psi_{\rm md}\rangle$ and this isometry is not evident in these geometries.

F. Entangled qubits and linearity

We now consider a final case in some detail: the situation where the CFT is entangled with a few qubits. In this situation not only is there no geometric wormhole, but we find that it is possible to select the interior operators to strictly commute (as operators) with all operators in the qubit system.

For now we make no assumption about the Hamiltonian of the qubit system. However, the combined CFT and qubit system can be in equilibrium only in states of the form

$$|\Psi_{\rm qub}\rangle = \sum_i \alpha_i |E_{qi}\rangle \otimes |\Psi_i\rangle, \tag{9.27}$$

where $|E_{qi}\rangle$ are energy eigenstates in the qubit system and $|\Psi_i\rangle$ are equilibrium states in the CFT, and the coefficients α_i obey $\sum_i |\alpha_i|^2 = 1$.

The reason that the entanglement structure has to be of this form in an equilibrium state is because in the qubit system, we assume that we have access to *all* operators. Therefore the only equilibrium states in this system are strict energy eigenstates which remain invariant under time evolution. If, upon tracing out the CFT, we were to obtain any significant off-diagonal terms in the qubit density matrix, then it would be possible to find an appropriate operator whose expectation value would be time dependent. These energy eigenstates must be entangled with states that are independently in equilibrium in the CFT. This fixes equilibrium states to be of the form (9.27).

We now find that

$$\mathcal{H}^0_{\Psi_{\rm qub}} = \sum_i \alpha_i |E_{qi}\rangle \otimes \mathcal{A}|\Psi_i\rangle.$$

We now act with an arbitrary operator from the qubit system $\boldsymbol{A}_{\rm L,1}$ to obtain

$$\boldsymbol{A}_{\rm L,1}|\Psi_{\rm qub}\rangle = \sum_{i,j} \alpha_i A^{ji}_{L,1} |E_{qj}\rangle \otimes |\Psi_i\rangle, \tag{9.28}$$

where $A^{ji}_{L,1}$ are the matrix elements in the qubit-energy eigenbasis of the left operator. This state is not in left equilibrium but because a small superposition of equilibrium state is still an equilibrium state we see that (9.28) still represents a right-equilibrium state and does not lie in $\mathcal{H}^0_{\Psi_{\rm qub}}$.

Proceeding in this manner, we find that the little Hilbert space has the form

$$\mathcal{H}_{\Psi_{\rm qub}} = \bigoplus_{i,j} |E_i\rangle \otimes \mathcal{A}|\Psi_j\rangle.$$

Now, using the prescription above, we find that the action of the mirrors is given by

$$\tilde{\boldsymbol{\mathcal{O}}}_{\omega,m}(|E_i\rangle \otimes \boldsymbol{A}_\alpha|\Psi_j\rangle) = |E_i\rangle \otimes \boldsymbol{A}_\alpha e^{-\frac{\beta\omega}{2}} \boldsymbol{\mathcal{O}}^\dagger_{\omega,m}|\Psi_j\rangle. \tag{9.29}$$

Therefore in this situation the mirror operators are entirely operators within the right CFT and do not act in the qubit system at all. Moreover the mirror operators above can be understood as follows. We construct mirror operators on each of the equilibrium states $|\Psi_i\rangle$. We then take the union of these operators and this yields the operators above.

Avoiding possible superluminality in the presence of state-dependence: Let us briefly mention the significance of the observation above. Our state-dependent operators are sometimes conflated with notions of "nonlinear" quantum mechanics that have been proposed earlier. Gisin [63] and Polchinski [64] pointed out sharp difficulties with one such idea that was advanced by Weinberg [65]. In particular, Gisin noted that nonlinear evolution in quantum mechanics could lead to superluminal communication.

We emphasize that in our proposal we do not add any nonlinear terms to the Hamiltonian, which is simply the CFT Hamiltonian. Nevertheless, one may still be concerned about this issue of superluminality. We now show that this also does not arise in our construction.

Consider the following experiment. An experimenter entangles black hole microstates in the CFT with states of a "small pointer" comprising a few qubits. Then the qubits and the CFT are separated by a large distance. An observer from the CFT now jumps into the black hole and makes a measurement. Physically, we expect that such an observer should not be able to send messages to another observer who has access only to the qubits.

To make this more precise, consider a qubit system with $M+1$ states, that we denote by $|1\rangle, |2\rangle, \ldots |M+1\rangle$, where $M \ll \mathcal{N}$. Now, we consider $\boldsymbol{M}$ equilibrium states of the CFT, $|\Psi_1\rangle \ldots |\Psi_M\rangle$, and take them to be orthogonal without loss of generality. Let us prepare the joint qubit-CFT system in the state

$$|\Psi_{\rm qub}\rangle = \sum_{i=1}^{M} \alpha_i |i\rangle \otimes |\Psi_i\rangle + |M+1\rangle \otimes \left(\sum_j \beta_j |\Psi_j\rangle\right). \tag{9.30}$$

In order for the state to be normalized correctly, we have the condition

$$\sum_i |\alpha_i|^2 + |\beta_i|^2 = 1.$$

Now, we act with a *unitary* of the mirror operators on $|\Psi_{\rm qub}\rangle$. Let us call this unitary $\tilde{\boldsymbol{U}}$. We see that from (9.29) we have

$$\tilde{\boldsymbol{U}}|\Psi\rangle = \sum_{i=1}^{M} \alpha_i |i\rangle \otimes \tilde{\boldsymbol{U}}|\Psi_i\rangle + |M+1\rangle \otimes \tilde{\boldsymbol{U}}\left(\sum_j \beta_j |\Psi_j\rangle\right). \tag{9.31}$$

The key physical requirement to ensure that no messages can be sent from the black hole interior to the qubit system is that this process should leave the density matrix of the pointer invariant. The density matrix of the pointer in (9.30) has the following components:

$$\begin{aligned} \langle M+1|\boldsymbol{\rho}^{\rm init}|M+1\rangle &= \sum |\beta_i|^2, \\ \langle i|\boldsymbol{\rho}^{\rm init}|i\rangle &= |\alpha_i|^2, \\ \langle i|\boldsymbol{\rho}^{\rm init}|M+1\rangle &= \alpha_i \beta_i^*, \\ \langle M+1|\boldsymbol{\rho}^{\rm init}|i\rangle &= \alpha_i^* \beta_i. \end{aligned} \tag{9.32}$$

For convenience, let us denote $|\chi\rangle = \tilde{\boldsymbol{U}}(\sum_j \beta_j |\Psi_j\rangle)$. Then the components of the density matrix of the pointer in the final state (9.31) are

$$\begin{aligned} \langle M+1|\boldsymbol{\rho}^{\rm fin}|M+1\rangle &= \langle\chi|\chi\rangle, \\ \langle i|\boldsymbol{\rho}^{\rm fin}|i\rangle &= |\alpha_i|^2, \\ \langle i|\boldsymbol{\rho}^{\rm fin}|M+1\rangle &= \alpha_i \langle\chi|\tilde{\boldsymbol{U}}|\Psi_i\rangle, \\ \langle M+1|\boldsymbol{\rho}^{\rm fin}|i\rangle &= \alpha_i^* \langle\Psi_i|\tilde{\boldsymbol{U}}^\dagger|\chi\rangle. \end{aligned} \tag{9.33}$$

Demanding that the infalling observer cannot send messages is equivalent to setting $\boldsymbol{\rho}^{\rm fin} = \boldsymbol{\rho}^{\rm init}$. From (9.32)–(9.33) we see that this implies

$$\begin{aligned} \langle\chi|\chi\rangle &= \sum |\beta_i|^2, \\ \langle\chi|\tilde{\boldsymbol{U}}|\Psi_i\rangle &= \beta_i^*, \\ \langle\Psi_i|\tilde{\boldsymbol{U}}^\dagger|\chi\rangle &= \beta_i. \end{aligned}$$

In fact, since the states $\tilde{\boldsymbol{U}}|\Psi_i\rangle$ also give an orthogonal set, we see that we are forced to the conclusion that

$$|\chi\rangle = \beta_i \tilde{\boldsymbol{U}}|\Psi_i\rangle.$$

This implies that the operator $\tilde{\boldsymbol{U}}$ must act linearly on a superposition of a small number of states.

This is precisely what is ensured by the construction above. As we mentioned, this construction proceeds by constructing mirrors for each of the individual equilibrium states and then just taking the union of their actions, which ensures that the constraint above is satisfied. The reader may recall the discussion of Sec. VII E where we verified that our operators naturally respect linearity in their action on small superpositions.

This result is important because it shows that in the context of entanglement with pointers, and experiments of the kind considered above, the state-dependence of our operators is *completely transparent* to the infalling observer. Therefore, in no experiment that can be described within effective field theory does the observer detect a violation of linearity.

We conclude by remarking on a slightly subtle point. We have now described two situations where there is entanglement but no geometric wormhole between the CFT and the system that it is entangled with. However, from the point of view of the microscopic operators, this is attained rather differently when the left system is a CFT, and when it is just a collection of qubits. In the case where the left system is a CFT and the entanglement entropy is large, the right mirror operators commute with simple left operators but not with all operators on the left. On the other hand, in the case where the CFT is entangled with a few qubits or with a system that does not have $O(e^{\mathcal{N}})$ states, then we can indeed find mirrors entirely within the original CFT. As we saw above this was important to ensure the absence of superluminal effects in such cases.

G. Refining the notion of equilibrium for entangled states

In some cases, the fact that our notion of equilibrium as time independence of simple correlators is necessary but not sufficient—as we discussed in Sec. VIII C—is also relevant to the discussion of entangled states. Consider the state

$$\boldsymbol{M}(\boldsymbol{A}_\alpha)|\Psi_{\rm en}\rangle = e^{-\frac{\beta \boldsymbol{H}}{2}} (e^{i\boldsymbol{A}_\alpha})^\dagger e^{\frac{\beta \boldsymbol{H}}{2}} |\Psi_{\rm en}\rangle. \tag{9.34}$$

In the thermofield state, correlation functions of this state are time invariant on the right, but not on the left. This is because we have

$$\boldsymbol{M}(\boldsymbol{A}_\alpha)|\Psi_{\rm tfd}\rangle = e^{i\boldsymbol{A}_{{\rm L},\alpha}}|\Psi_{\rm tfd}\rangle.$$

Therefore, in this case, this lack of equilibrium can be detected by our left-equilibrium criterion.

On the other hand, in a generic entangled state there is no such relation between these states and left-excited states. Therefore, in such states the ambiguity from the single-sided case carries over. The reason we imposed the restriction that the left excitation in (9.6) be Hermitian was to prevent this ambiguity in descendants. Given the state in (9.6) we can dress it with a left unitary to obtain another valid descendant, which also appears to be in right equilibrium. With $\boldsymbol{A}^U_{{\rm L},1} = e^{i\boldsymbol{A}_{{\rm L},\alpha}} \boldsymbol{A}_{{\rm L},1}$, we could have considered

$$|\Psi_{\rm en}^{1,U}\rangle = (1 - \boldsymbol{P}_{\rm en}^{\boldsymbol{0}})\boldsymbol{A}_{\rm L,1}^{U}|\Psi_{\rm en}\rangle$$

in (9.7). However, when $\boldsymbol{A}_{\rm L,\alpha}$ is entangled with a right operator, we want to ensure that we do not mistake $|\Psi_{\rm en}^{1,U}\rangle$ for an equilibrium descendant. However, the restriction that the left excitation be Hermitian excludes operators of the form $\boldsymbol{A}_{\rm L,1}^{U}$.

As we explained in Sec. VIII C, even though all correlators on the right are left invariant under the excitation (9.34), it should still be possible to find measurables that can detect this excitation. Although we have not yet identified such measurables precisely, it is possible that the physical quantity that is capable of detecting the excitation in (8.21) in a single-sided CFT will also be able to detect the excitation (9.34) in the two-sided case.

X. DISCUSSION

In this paper we have presented strong evidence for the claim that the black hole interior must be described using state-dependent bulk-boundary maps. We showed that a state-independent construction of the interior was impossible, not only for single-sided AdS black holes, but even for the eternal black hole. It is possible that this indicates that AdS/CFT does not describe black hole interiors at all. However, this is in contradiction with many other calculations that suggest that the eternal black hole, at least, does have a smooth interior that can be probed by the CFT.

State-dependent bulk to boundary maps provide a solution to these versions of the information paradox that preserves the predictions of effective field theory. Our state-dependent construction of the black hole interior explicitly identifies the duals of bulk local operators in the CFT. These bulk probes do not see any sign of a pathology at the horizon, and so this should be taken as additional evidence that generic states do not correspond to firewalls.

In this paper, we demonstrated that our construction does not lead to any violation of quantum mechanics or the Born rule. We also successfully resolved some of the ambiguities in our definition of an equilibrium state.

Furthermore, we showed that our construction admitted a natural extension to entangled systems. This extension leads to a surprising bonus: a precise version of the ER = EPR conjecture emerges automatically from our construction without having to put anything in by hand.

We have described our construction in significant detail and discussed how it works in equilibrium states—which are generic at high energy. We have also considered a large class of nonequilibrium states, including those that have been excited outside and inside the horizon. Although it is possible to consider other special classes of states in the CFT, we believe that our results provide persuasive evidence for the consistency of our construction.

There are several natural questions that arise from this analysis. It would be interesting to examine local operators outside the horizon in greater detail. Although we presented a state-independent description of such operators, in the minisuperspace approximation in Sec. IV B 2, the question of whether state-dependence is also required outside the horizon is open. We comment more on this in [39].

It would also be interesting to understand whether our construction can shed some light on the nature of the black hole singularity. So far we have used techniques from effective field theory to motivate the bulk to boundary map. Any investigation of the singularity requires new ideas.

Recent studies [66] have shown that the naive $\frac{1}{N}$ expansion can often break down unexpectedly. We would like to understand the implications of this breakdown for effective field theory on the nice slices and for the limitations of locality in quantum gravity.

Finally, as we have explained, while the use of state-dependent operators is perfectly consistent with quantum effective field theory, they are both unusual and interesting. It would be very useful to develop a more comprehensive measurement theory for these objects and understand whether they appear in other settings.

ACKNOWLEDGMENTS

We have discussed these ideas with a large number of people over the past year. We are particularly grateful to all members of ICTS-TIFR, and the string theory groups at TIFR (Mumbai), IISc (Bangalore), CERN, the University of Groningen and Harvard University. We are also grateful to the organizers and participants of the "Bulk Microscopy from Holography and Quantum Information" (PCTS, Princeton, 2013); the summer workshop on emergent spacetime in string theory (2014) at the Aspen Center for Physics (which was supported by NSF Grant No. PHYS-1066293 and the Simons Foundation); the Santa Barbara Gravity Workshop (2014); the CERN Winter School on Strings and Supergravity (2015); the QUC Autumn Symposium on String/M Theory (KIAS, Seoul, 2014); the IPM Winter School and Workshop (Tehran, 2014); the Asian Winter School (Puri, 2014); the discussion meeting on the black hole information paradox (HRI, Allahabad, 2014); the discussion meeting on entanglement and gravity (ICTS-TIFR, 2014); the Bangalore Area Discussion Meeting (ICTS-TIFR, 2015); the "workshop on holography, gauge theory and black holes" (Southampton, 2014); the "Solvay Workshop on Holography for Black Holes and Cosmology" (Brussels, 2014); the COST meeting "String theory Universe" (Mainz, 2014); the "Institut d'ete" (ENS, 2014); the "Amsterdam Summer String Workshop" (2014); the Nordic String Meeting 2015 (Groningen) and the Swiss String Meeting, GeNeZiSS 2015 (Bern). K. P. acknowledges the hospitality of the Institute for Advanced Study (Princeton) and of the Crete Center for Theoretical Physics. S. R. also acknowledges the hospitality of the Center for the Fundamental Laws of Nature at Harvard, Delhi University,

and the Institute for Advanced Study (Princeton). S. R. is partially supported by a Ramanujan fellowship of the Department of Science and Technology (India). S. R. also acknowledges the partial support of the Center of Mathematical Sciences and Applications at Harvard University and the Cheng Yu-Tung Fund for Research at the Interface of Mathematics and Physics. K. P. thanks the Royal Netherlands Academy of Sciences (KNAW).

APPENDIX A: STATE-DEPENDENCE AND SEMICLASSICAL QUANTIZATION

In this appendix, we explore the semiclassical origins of state-dependence. Some of the ideas in this appendix were anticipated in [14], although our analysis differs in some eventual details. As we mentioned in Sec. III, the belief that geometric quantities such as the metric should be represented by state-independent operators in the CFT is predicated on intuition from geometric quantization. We elaborate on this intuition here. But we also explain why this intuition fails because of important ways in which the Hilbert space of the CFT differs from what one might expect from a semiclassical linearized analysis of gravity.

1. Review of semiclassical quantization

We briefly remind the reader of the elementary concepts involved in quantizing the phase space of a system so as to make the classical limit manifest. We closely follow the excellent review by Yaffe [67].

Before we proceed to the analysis for gravity, we briefly remind the reader of the elementary notions that are involved in semiclassical quantization. Consider a system with canonical variables x_i, p_i, with $i = 1 \ldots n$, obeying the classical Poisson bracket relations $\{x_i, p_i\}_{\text{P.B.}} = 1$, and some classical functions on the phase space $f_m(\vec{x}, \vec{p})$. We assume that all the first class constraints have been converted to second class constraints by gauge fixing and that all the second class constraints have been solved to eliminate the dependent variables. So the phase space is unconstrained.

Here we have denoted the coordinates on phase space by two vectors $\vec{x}, \vec{p}$, with $\vec{x} = (x_1, \ldots x_n)$ and $\vec{p} = (p_1, \ldots p_n)$. We also define $\vec{z} = (\frac{1}{\sqrt{2}}(x_1 + ip_1), \ldots \frac{1}{\sqrt{2}}(x_n + ip_n))$. Now, we want to show that in the quantum theory it is possible to find (a) an appropriate set of operators $\hat{f}_m$ and (b) a set of semiclassical *coherent* states $|\vec{x}, \vec{p}\rangle$ in one to one correspondence with the phase space so that, *when evaluated* on these states the operators $\hat{f}_m$ behave like the classical functions $f_m(x, p)$ as we discuss more precisely below.

First, since we already have a simple and explicit description of the phase space and symplectic form in this setting, we quantize the system and define the canonical operators $\hat{x}_i, \hat{p}_i$ satisfying $[\hat{x}_i, \hat{p}_j] = i\delta_{ij}$. This provides us with eigenstates of the operators $\hat{x}_i$ that satisfy $\hat{x}_i|\vec{x}\rangle = x_i|x_1, \ldots x_n\rangle$. We also define $\hat{a}_i = \frac{1}{\sqrt{2}}(\hat{x}_i + i\hat{p}_i); \hat{a}_i^\dagger = \frac{1}{\sqrt{2}}(\hat{x}_i - i\hat{p}_i)$.

With the vacuum $|\Omega\rangle$ defined as $a_i|\Omega\rangle = 0$, we consider the coherent states

$$|\vec{z}\rangle = e^{-\frac{\sum_i |z_i|^2}{2}} e^{\sum_i a_i^\dagger z_i}|\Omega\rangle.$$

The wave function of this state in the basis of eigenvectors of $\hat{x}_i$ can be calculated by noticing that $a_i|\vec{z}\rangle = z_i|\vec{z}\rangle$. With $\Psi_z(\vec{x}_i) = \langle\vec{x}|\vec{z}\rangle$, and using the fact that in the position eigenbasis $\hat{p}_i = -i\frac{\partial}{\partial x_i}$, this turns into the differential equation

$$\left(x_i + \frac{\partial}{\partial x_i}\right)\Psi_{\vec{z}}(\vec{x}) = (z_{xi} + iz_{pi})\Psi_{\vec{z}}(\vec{x}),$$

where we have written the components of z_i as $z_i = z_{xi} + iz_{pi}$ to avoid confusion with the x_i variable on the left. This is solved by the normalized position space wave function for the coherent states.

$$\Psi_{\vec{z}}(\vec{x}) = \left(\frac{2}{\pi}\right)^{\frac{n}{4}} \exp\left\{-\sum_i [(x_i - z_{xi})^2 + iz_{pi}(x_i - z_{xi})]\right\}. \tag{A1}$$

These states play the role of semiclassical states, and we can place them in a bijective correspondence with the phase space.

These coherent states have several important properties. They are not orthonormal; in fact, it is important that they form an overcomplete basis of the Hilbert space. We have

$$\langle\vec{u}|\vec{z}\rangle = e^{-\frac{|\vec{z}|^2}{2}} e^{-\frac{|\vec{u}|^2}{2}} \langle\Omega|e^{\vec{a}\cdot\vec{u}} e^{\vec{a}^\dagger\cdot\vec{z}}|\Omega\rangle = e^{-\frac{|\vec{z}|^2}{2} - \frac{|\vec{u}|^2}{2} + \vec{u}\cdot\vec{z}},$$
$$|\langle\vec{u}|\vec{z}\rangle|^2 = e^{-|\vec{z}-\vec{u}|^2}. \tag{A2}$$

Nevertheless, we can partition the identity by using projectors onto these states.

$$1 = \frac{1}{(2\pi)^n}\int d^{2\vec{}}\vec{z} P_{\vec{z}}; \qquad P_{\vec{z}} = |\vec{z}\rangle\langle\vec{z}|. \tag{A3}$$

This identity can be easily proved using, for example, the position space representation of the coherent states in (A1).

Next, we need a way of lifting functions from the phase space to operators. Consider a function $f(\vec{z})$ on the phase space. (We have suppressed the dependence on $\vec{\bar{z}}$ simply to lighten the notation; we do not necessarily consider only holomorphic functions.) We now consider the operator defined by

$$\hat{f} = \int f(\vec{z})|\vec{z}\rangle\langle\vec{z}|\frac{d^{2n}\vec{z}}{(2\pi)^n}. \tag{A4}$$

This representation of operators is the so-called Sudarshan-Mehta P-representation [68]. It differs from the more commonly used Weyl representation of operators, by operator ordering. The Weyl representation is sometimes favored in the literature, since this map also allows one to represent the product of operators in the quantum theory by a Moyal star product of functions on the phase space. However (A4) yields more insight for our discussion, and has the same classical limit as the Weyl representation.

Note that when this operator is inserted back into a coherent state we have

$$\langle \vec{u}|\hat{f}|\vec{u}\rangle = \int f(\vec{z}) e^{-|\vec{z}-\vec{u}|^2} \frac{d^{2n}\vec{z}}{(2\pi)^n}.$$

Therefore, the expectation value of the quantum operator is a slightly smeared version of the classical function. We have suppressed factors of $\hbar$ here, but if we consider classical functions that do not vary rapidly within a volume of $\hbar$ about a point in phase space, then the expectation value of the corresponding quantum operators faithfully reproduces their behavior.

Furthermore, if we consider the expectation value of the product of two operators then by using (A3)

$$\begin{aligned}\langle \vec{y}|\hat{f}\,\hat{g}\,|\vec{y}\rangle &= \frac{1}{(2\pi)^2}\int f(\vec{z})g(\vec{u})\langle \vec{y}|\vec{u}\rangle\langle \vec{u}|\vec{z}\rangle\langle \vec{z}|\vec{y}\rangle d^2\vec{z}d^2\vec{u} \\ &= \frac{1}{(2\pi)^2}\int f(\vec{z})g(\vec{u})e^{-|\vec{z}|^2-|\vec{u}|^2-|\vec{y}|^2+\vec{u}\cdot\vec{z}+\vec{y}\cdot\vec{u}+\vec{z}\cdot\vec{y}} \\ &\quad \times d^2\vec{z}d^2\vec{u}.\end{aligned}$$

We see that this integral is peaked around $z = u = y$ and expanding $g(\vec{u}) = g(\vec{y}) + (\vec{u}-\vec{y})\cdot\partial_{\vec{y}}g(\vec{y}) + \ldots$, and similarly for f, we see that the leading term is obtained by doing the Gaussian integral and we find

$$\langle \vec{y}|\hat{f}\,\hat{g}\,|\vec{y}\rangle \approx f(\vec{y})g(\vec{y}).$$

On the other hand, we can also compute the commutator between two functions, in which case we need to keep the first subleading term to obtain a nonzero answer. Here, we find

$$\langle \vec{y}|[\hat{f},\hat{g}]|\vec{y}\rangle = i\{f,g\}_{P.B.}(\vec{y}).$$

2. Geometrical quantities as classical functions on the phase space

We now turn to the case of gravity where we first discuss the classical phase space and then describe coherent states in the linearized theory. In this subsection we are interested in establishing the following

Claim: "the metric $g_{\mu\nu}(\vec{x})$ is a well-defined function on the classical phase space of gravity."

The phase space of gravity is often discussed in canonical terms, where we specify the three-metric and the extrinsic curvature on a spacelike slice. This provides Cauchy data that we can evolve forward and backward in time. However, a covariant description of the phase space is given by considering the set of all classical solutions to gravity with asymptotic AdS boundary conditions [69–71]. The map between these two pictures is straightforward.

Given a solution to the classical equations of motion, and a metric with a $d+1$ split,

$$ds^2 = -N^2dt^2 + \gamma_{ij}(dx_i + N_i dt)(dx_j + N_j dt), \quad \text{(A5)}$$

one may simply evaluate the fields at the spacelike slice $t = 0$. Then the variables,

$$\gamma_{ij}(\vec{x},0), \qquad \pi^{ij}(\vec{x},0) = -\gamma^{\frac{1}{2}}(K^{ij} - \gamma^{ij}K),$$

provide the standard parametrization of gravitational phase space. Here K is the extrinsic curvature

$$K_{ij} = \frac{1}{2}N^{-1}(\partial_j N_i + \partial_i N_j - \partial_t \gamma_{ij}), \quad \text{(A6)}$$

and for the purposes of this $d+1$ split we have displayed the time coordinate separately in $(\vec{x},t)$.

Conversely, given the variables $\gamma_{ij}(\vec{x},0)$ and $\pi^{ij}(\vec{x},0)$, one may use the equations of motion to evolve them forward in time and generate the entire metric in the form (A5). Of course, such a solution requires a choice of gauge, as we have already discussed.

It is also possible to write down a symplectic form on the phase space described covariantly as the set of classical solutions, and this was done by [70].

For us the important point is that each point on the phase space corresponds to an entire spacetime. Now, evidently given the entire spacetime, classically, we may ask any question we wish, even one that involves global notions like an event horizon. For example, we may set up relational coordinates as in Sec. III A 1 and just evaluate the metric at a point $g_{\mu\nu}(\vec{x},t)$. The same is true of other propagating light fields in the theory.

Therefore, all of these observables are well-defined classical functions on the phase space. This is an important point. We now extend the discussion above to gravity to show that, explicitly, within the linearized theory, we may indeed expect such questions to be answered by state-independent operators.

3. Coherent states in linearized gravity

We now turn to an analysis of gravity. Here we are interested in establishing the following.

Claim: If we consider two nearby points in the gravitational phase space with metrics $g^{b}_{\mu\nu}(\vec{x})$ and $g^{e}_{\mu\nu}(\vec{x})$ then

one can define a covariant inner product on the corresponding coherent states in the Hilbert space which behaves like $e^{-\mathcal{N}\upsilon(g^b,g^e)}$ where we can compute the function υ in the linearized approximation.

First we remind the reader how the discussion of A 1 generalizes to linearized gravity. We are only able to work in the linearized setting, and although it would be interesting to explore this construction further in a fully nonlinear setting, we do not know how to do this.

We consider fluctuations of the metric, about a background metric, defined by

$$g_{\mu\nu} = g^b_{\mu\nu} + \sqrt{8\pi G_N} h_{\mu\nu},$$

and the normalization is chosen so that the kinetic term of $h_{\mu\nu}$ is canonically normalized. Here $g^b_{\mu\nu}$ may be any background metric that is a solution of the equations of motion and is asymptotically AdS. We do not take it to be necessarily the AdS-Schwarzschild solution.

Now, on general grounds, we expect that solutions to the classical equations of motion will be given by

$$h_{\mu\nu}(\vec{x}) = \sum_{i,\omega} a^i_\omega g^{(i)}_{\mu\nu}(\omega, \vec{x}) + \text{H.c.},$$

where i runs over the different $\frac{(d+1)(d-2)}{2}$ possible polarizations of the graviton, where d is the boundary dimension and a^i_ω are just linear coefficients at the moment. The different eigenfunctions are denoted by ω. In empty AdS or AdS Schwarzschild, for example, this would constitute a set of integers to pick out the spherical harmonic on the S^{d-1} and a "radial momentum." We do not require the detailed form of these eigenfunctions, or even of their eigenvalues. We are not assuming that there is a timelike isometry in the space, and so, in principle, ω may not correspond intuitively to a frequency.

We also assume that we have picked a basis set of distinct solutions $g^{(i)}_{\mu\nu}$, which are not equivalent under gauge transformations, and we normalize the functions $g^{(i)}_{\mu\nu}(\omega, \vec{x})$ so that the canonical Poisson brackets translate into the statement

$$\{a^i_\omega, a^{j,\dagger}_{\omega'}\}_{P.B.} = -i\delta^{ij}\delta_{\omega,\omega'}.$$

We quantize the theory and obtain a vacuum state $a^i_\omega|\Omega\rangle = 0$. Note that now a^i_ω is an operator on the Hilbert space of the linearized theory. We then define coherent states by labeling them with a set of functions $\chi^i(\omega)$. Starting with the vacuum,

$$|\chi\rangle \equiv \mathcal{N}_\chi e^{\sum_{i,\omega} a^{i,\dagger}_\omega \chi^i_\omega}|\Omega\rangle,$$

where $\mathcal{N}_\chi$ is a normalization factor. We see that

$$\langle\chi|\chi\rangle = |\mathcal{N}_\chi|^2 e^{\sum_{i,\omega}|\chi^i_\omega|^2}.$$

So for the state to be normalized, we should set

$$\mathcal{N}_\chi = e^{-\frac{1}{2}\sum_{i,\omega}|\chi^i_\omega|^2}. \tag{A7}$$

Note that $|\chi^i_\omega|^2$ can also be interpreted as the "occupation number" in the mode ω; so the exponent in the normalization factor is just the total occupation number in the state.

One measure of how large the deviation of the field is from the background metric is given by

$$\langle\Omega|\chi\rangle = \mathcal{N}_\chi. \tag{A8}$$

Here the vacuum is just the original background metric. So we see that this coherent state is substantially different from the original background metric, as a quantum state, if the occupation number is large. In this state the metric has an expectation value

$$\begin{aligned} g^e_{\mu\nu} = \langle\chi|g_{\mu\nu}(\vec{x})|\chi\rangle &= \langle\chi|g^b_{\mu\nu}(\vec{x}) + \sqrt{8\pi G_N} h_{\mu\nu}(\vec{x})|\chi\rangle \\ &= g^b_{\mu\nu}(\vec{x}) + \sqrt{8\pi G_N} \\ &\quad \times \sum_{i,\omega}(\chi^i_\omega g^{(i)}_{\mu\nu}(\omega, \vec{x}) + \text{H.c.}). \end{aligned} \tag{A9}$$

So we see that the space $|\chi\rangle$ represents a nearby point in phase space, where the value of the metric has changed to $g^e_{\mu\nu}(\vec{x})$. Therefore (A7) shows how the corresponding inner product in Hilbert space varies.

Now, in deriving (A7) we made explicit reference to a set of mode functions. But we would like it to depend only on the two metrics $g^e_{\mu\nu}(\vec{x})$ and $g^b_{\mu\nu}(\vec{x})$. To check that this is covariant, let us consider how this changes under a Bogoliubov transformation of the modes. We make a canonical transformation of the a^i_ω variables to

$$\begin{aligned} b^i_\omega &= \sum_{\omega'}(\beta_{\omega\omega'} a^i_{\omega'} + \gamma_{\omega,\omega'} a^{\dagger,i}_{\omega'}), \\ b^{\dagger,i}_\omega &= \sum_{\omega'}(\beta^*_{\omega\omega'} a^{\dagger,i}_{\omega'} + \gamma^*_{\omega,\omega'} a^i_{\omega'}). \end{aligned} \tag{A10}$$

In this analysis, we assume that the polarization index i does not enter the Bogoliubov coefficients. This is just to lighten the notation and does not represent any loss of generality.

For the new modes to have the canonical commutators

$$[b^i_\omega, b^{\dagger,i}_{\omega'}] = \delta_{\omega,\omega'},$$

we see that we must have

$$\sum_{\omega''}(\beta_{\omega,\omega''}\beta^*_{\omega',\omega''} - \gamma_{\omega,\omega''}\gamma^*_{\omega',\omega''}) = \delta_{\omega,\omega'}. \tag{A11}$$

An observer using these creation and annihilation operators would also use a new basis of modes to represent the metric fluctuations that we call $\tilde{g}^{(i)}(\omega,\vec{x})$. In particular, we have

$$\sum_{\omega}\beta_{\omega\omega'}\tilde{g}^{(i)}(\omega,\vec{x})+\gamma^{*}_{\omega,\omega'}(\tilde{g}^{(i)}(\omega,\vec{x}))^{*}=g^{(i)}(\omega',\vec{x}),$$
$$\sum_{\omega}\beta^{*}_{\omega\omega'}(\tilde{g}^{(i)}(\omega,\vec{x}))^{*}+\gamma_{\omega,\omega'}(\tilde{g}^{(i)}(\omega,\vec{x}))=(g^{(i)}(\omega',\vec{x}))^{*}. \tag{A12}$$

Such an observer would set up a different set of coherent states

$$|\tilde{\chi}\rangle_{\mathrm{Bog}}=e^{\tilde{\chi}^{i}_{\omega}b^{\dagger,i}_{\omega}}|\Omega\rangle_{\mathrm{Bog}},$$

where the vacuum is now defined to satisfy $b^{i}_{\omega}|\Omega\rangle_{\mathrm{Bog}}=0$. To get the same expectation value for the metric field, this observer could use a coherent state excitation with parameters $\tilde{\chi}^{i}_{\omega}$ so that

$$\sum_{\omega}\tilde{\chi}^{i}_{\omega}\tilde{g}^{i}(\omega,\vec{x})+(\tilde{\chi}^{i}_{\omega})^{*}(\tilde{g}^{i}(\omega,\vec{x}))^{*}=\sum_{\omega'}\chi^{i}_{\omega'}g^{(i)}(\omega',\vec{x})+(\chi^{i}_{\omega})^{*}g^{(i)}(\omega',\vec{x}).$$

Using (A12), we see that we need

$$\tilde{\chi}^{i}_{\omega}=\sum_{\omega'}(\beta_{\omega\omega'}\chi^{i}_{\omega'}+\gamma_{\omega,\omega'}(\chi^{i}_{\omega'})^{*}).$$

Therefore we see that

$$\sum_{i,\omega}|\tilde{\chi}^{i}_{\omega}|^{2}=\sum_{i,\omega,\omega',\omega''}[\beta_{\omega\omega'}\beta^{*}_{\omega\omega''}\chi^{i}_{\omega'}(\chi^{i}_{\omega''})^{*}+\gamma_{\omega,\omega'}\gamma^{*}_{\omega,\omega''}\chi^{i}_{\omega''}(\chi^{i}_{\omega'})^{*}+\beta_{\omega\omega'}\gamma_{\omega\omega''}\chi^{i}_{\omega}\chi^{i}_{\omega''}+\beta^{*}_{\omega\omega'}\gamma^{*}_{\omega\omega''}(\chi^{i}_{\omega'})^{*}(\chi^{i}_{\omega''})^{*}]. \tag{A13}$$

For a general Bogoliubov transformation therefore

$$\sum_{i,\omega}|\tilde{\chi}^{i}_{\omega}|^{2}=\sum_{i,\omega}|\chi^{i}_{\omega}|^{2}+R, \tag{A14}$$

where the remainder R does not vanish.

However, in AdS/CFT we have an additional advantage: the presence of the boundary Hamiltonian. So we can define positive and negative energy with respect to the boundary Hamiltonian and demand that in terms of boundary energy eigenstates, both the sets of creation operators have strictly positive energy and the annihilation operators have negative energy.[27]

[27]Here, we are not concerned with the small tails that we discussed in the text, which may appear in these relations because we restrict observations to a finite time on the boundary.

$$P_{E^{+}}a^{i}_{\omega}|E\rangle=0,\qquad P_{E^{+}}b^{i}_{\omega}|E\rangle=0,$$
$$P_{E^{-}}a^{i,\dagger}_{\omega}|E\rangle=0,\qquad P_{E^{-}}b^{i,\dagger}_{\omega}|E\rangle=0, \tag{A15}$$

where $P_{E^{+}}$ ($P_{E^{-}}$) indicates the projector on the subspace formed by eigenstates with energy larger (smaller) than E. If we restrict to such operators then we see that $\gamma_{\omega\omega'}$ in (A10) must vanish. From (A11), we then find that $\beta_{\omega\omega'}$ must be *unitary*. For this set of transformations, which obeys the natural AdS/CFT constraint (A15), we see from (A13) that $R=0$ in (A14).

To summarize, the conclusion is that using the AdS/CFT Hamiltonian to define positive energy, the notion of the distance of a coherent excitation from the background is robust in linearized gravity.

Now, let us examine this distance a little more closely. Let us write the initial metric in a nice coordinate system so that all its components are of order the AdS radius squared ℓ^{2}. In this case, we see that to make a substantial perturbation, we must take $h_{\mu\nu}\sim\frac{\alpha\ell^{2}}{\sqrt{8\pi G_{N}}}=\alpha\mathcal{N}$, where α is an O(1) parameter that we have introduced. At this point, the linearized theory is still valid if we keep $\alpha\ll 1$. If we apply (A8) to such a perturbation, we see that the coherent state construction predicts the following. The semiclassical states in the quantum theory, corresponding to two distinct solutions $g^{e}_{\mu\nu}(\vec{x})$ and $g^{b}_{\mu\nu}(\vec{x})$, are almost orthogonal and have an inner product

$$\langle g^{e}_{\mu\nu}(\vec{x})|g^{b}_{\mu\nu}(\vec{x})\rangle=e^{-\mathcal{N}v(g^{e},g^{b})}, \tag{A16}$$

where v is a smooth O(1) functional on the space of metrics. To compute this function, we write $g^{e}_{\mu\nu}(\vec{x})$ as an excitation over $g^{b}_{\mu\nu}(\vec{x})$ using (A9) and compute the inner product given in (A7)–(A8). The choice of mode functions that we use to express the excited state in terms of the background is unimportant by the argument above.

a. Difficulties with state-independent operators

Now the formula for the inner product (A16) above might seem encouraging. It may suggest the following naive program. In the full theory of quantum gravity, we identify points on the phase space with coherent states $|g\rangle$, write down a completeness relation analogous to (A3) and then write a full state-independent metric operator as in (3.28): $\boldsymbol{g}_{\mu\nu}(\vec{x})=\sum_{g}g_{\mu\nu}(\vec{x})|g\rangle\langle g|$. This is the basis for the expectation that we can find state-independent operators to represent the metric and other bulk fields.

However, recall that (A3) was consistent only because the inner product (A2) died off to arbitrarily small values to compensate for the infinite volume of phase space. It appears that this does not happen for the case of gravity: rather, intuition from the CFT suggests that in some cases the inner-product between different coherent states may

saturate at a small but finite value even when the corresponding volume in classical phase space is very large.

We have seen an example of this in the case of the thermofield double. There the states $|\Psi_T\rangle$ all represented metrically distinct geometries. If we identify these states with points on the phase space, then the parameter T parametrizes an infinite direction in the classical phase space. However, even if we take T to be large, the inner product saturates at $\langle\Psi_{\rm tfd}|\Psi_T\rangle = O[e^{-\frac{S}{2}}]$ where S is the entropy.

This suggests that the classical limit in AdS/CFT emerges somewhat differently than the intuition from canonical gravity would suggest. Specifically, the following phenomenon occurs. We can identify states in the CFT dual to metrics $|\Psi_g\rangle \leftrightarrow |g\rangle$. However, when the distance between these states becomes large, the inner product in the CFT differs from the inner product predicted by semiclassical gravity. We have only been able to compute this semiclassical inner product reliably for small separations on the phase space. If we extrapolate this to the entire phase space then we can find cases where the semiclassical inner product is exponentially different from the CFT inner product.

$$\frac{e^{-\mathcal{N}v(g^a,g^b)}}{|\langle\Psi_{g^a}|\Psi_{g^b}\rangle|} = O(e^{-\mathcal{N}}).$$

Returning to the example of the thermofield double, which is the source of our intuition, we note that the formula (7.24) is precisely analogous to (A4). In both cases we know the action of an operator on a set of states that are almost orthogonal to one another. However, while in (A4) we are able to extend the integral to all of phase space and thereby obtain a state-independent operator; we cannot extend the limits on T in (7.24) to $\pm\infty$ because of the saturation of the inner product.

Another manifestation of this obstacle is as follows. In the thermofield double, given a sequence e^S states shifted by $\{T_1 \ldots T_{e^S}\}$, so that all of them are pairwise distinct, we can still find coefficients α_i so that

$$\left||\Psi_{\rm tfd}\rangle - \sum_{i=1}^{e^S} \alpha_i e^{i\boldsymbol{H}_L T_i}|\Psi_{\rm tfd}\rangle\right|^2 = O(e^{-\mathcal{N}}). \qquad \text{(A17)}$$

Note that (A17) is *not* due to Poincare recurrence, which occurs after a much longer time scale e^{e^S}. The linear dependence indicated in (A17) means that one geometry can be written as a linear combination of e^S completely different geometries. The semiclassical theory does not see any signs of (A17). This prevents a naive use of projectors on coherent states to build up a state-independent operator.

Summary The picture that we get in this manner is shown in Fig. 15. A slogan that would summarize this appendix is that "coherent states are always overcomplete, but the states in the CFT that correspond to coherent states of the metric are even more overcomplete than one would expect from a semiclassical analysis." This is what prevents us from lifting some well-defined classical observables to state-independent operators. This issue is important and interesting and deserves further investigation.

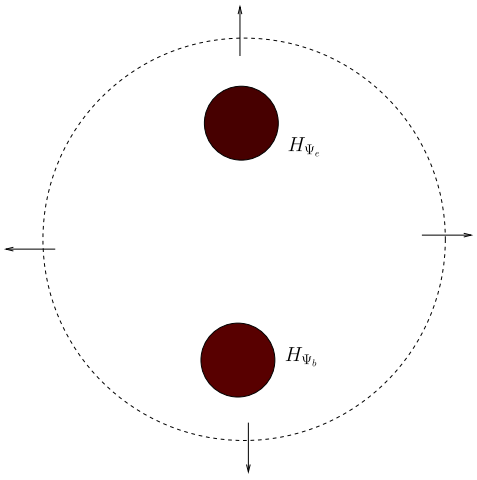


FIG. 15. When we quantize the theory we can put states in the Hilbert space in correspondence with the classical phase space. However, we may have to use different operators in different regions of phase space to represent a single classical function.

APPENDIX B: MIRROR MODES FROM BULK EVOLUTION

One possible proposal to define the mirror operators may proceed as follows. Consider black holes formed by collapse in AdS. In each such classical solution, we can trace the right moving modes behind the horizon to their origin to their support on the boundary of AdS in the past. This is what was done by Hawking in flat space [5] using a geometric optics approximation.

Hawking's computation suffers from a trans-Planckian problem because the geometric optics calculation tells us that, at late times, even low frequency right-moving modes behind the horizon come from an extremely small time band on the boundary. (See Fig. 16.) Therefore, in the past these low frequency modes must have had ultra-Planckian frequencies.

Even if we ignore this issue and proceed with the naive calculation, we find that we can only attain a small number of microstates by considering black holes formed from collapse. Page and Phillips estimated the number of possible configurations of massless radiation inside anti–de Sitter space [72]. Their calculation can be summarized as follows. Consider a gas of radiation in AdS_{d+1} and, as usual, set its radius to 1. Then, Page and Phillips considered a self-gravitating gas of radiation assuming that it was locally in thermal equilibrium at all points. Their conclusion was that one recovers the standard thermodynamic

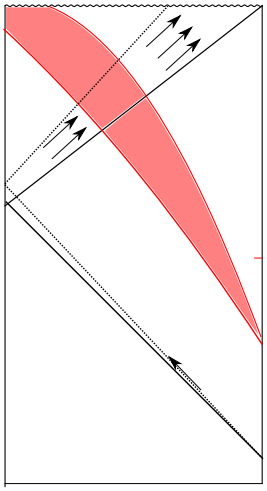


FIG. 16. Tracing the mirrors back to their origin on the boundary is difficult because of the trans-Planckian problem. However, even neglecting this issue does not help in constructing state-independent operators because of the fat tail in the inner product of different solutions.

relation between the entropy and the energy at high energies for a gas in $d+1$ dimensions,

$$S_{\rm rad} = \kappa_{\rm rad} E^{\frac{d}{d+1}}, \tag{B1}$$

where $\kappa_{\rm rad}$ is an O(1) constant which depends on the number of light degrees of freedom in the theory. On the other hand for high energies $E \gg \mathcal{N}$, we know that the entropy of the black hole is given by

$$S_{\rm bh} = \mathcal{N}\kappa_{\rm bh}\left(\frac{E}{\mathcal{N}}\right)^{\frac{d-1}{d}}, \tag{B2}$$

which is the result for a gas with $\mathcal{N}$ degrees of freedom in d dimensions. We remind the reader that $\mathcal{N}$ is the central charge, and so $\mathcal{N} = N^2$ in the SU(N) supersymmetric Yang-Mills theory.

Comparing (B2) with (B1) for energies of order $E \propto \mathcal{N}$, we find that

$$\frac{S_{\rm bh}}{S_{\rm rad}} = \frac{\kappa_{\rm bh}}{\kappa_{\rm rad}} \mathcal{N}^{\frac{1}{d}} E^{\frac{-1}{d(d+1)}} \propto \mathcal{N}^{\frac{1}{d+1}}.$$

Therefore the entropy of the radiation is always subleading in this range.

We caution the reader that (B1) is a little artificial in the regime in which we have applied it because the temperature that follows from (B1) is

$$T_{\rm rad} = \frac{1}{\left(\frac{\partial S_{\rm rad}}{\partial E}\right)} = \kappa_{\rm rad}^{-1} E^{\frac{1}{d+1}}.$$

If we consider the case of the duality between AdS_5 and supersymmetric Yang-Mills theory, with a 't Hooft coupling λ, then we do not expect the result (B1) to be valid beyond the string scale $\lambda^{\frac{1}{4}}$, at which point we expect to find a Hagedorn transition in the bulk. So, in reality we do not even expect to be able to attain as many microstates as we considered above for the radiating star.

This is a rather robust result: following the collapse of black holes from reasonable geometric configurations allows us to explore only a small fraction of the Hilbert space at high energies. Now if we do decide to restrict to such a sector of the Hilbert space, the firewall paradoxes vanish since they can only make reference to generic states. Correspondingly, there is no difficulty in obtaining state-independent mirror operators that have the correct behavior on this sector.

We now note a second important point. In some cases, it may be possible to geometrize the microstates of the black hole as we did in Sec. VI. There, we were able to explore a significant fraction of the microstates of the eternal black hole classically by considering a one-parameter family of eternal black hole solutions. All of these were glued to the boundary with different time shifts, and we had to allow this time shift to be exponentially large to ensure that the corresponding states in the CFT Hilbert space spanned a subspace of exponentially large dimension.

However, in this situation we ran into the obstruction explored in Sec. VII F and also in Appendix A. This obstacle is as follows. Any method of obtaining the mirror modes by analyzing classical solutions can, at most, specify these modes as functions on the classical phase space. For example in Sec. VII F, in each solution left shifted by the time T, the mirrors were the modes of $\mathcal{O}_{Lt+T,\Omega}$. However, in this situation we encountered the fat tail of (7.25). This fat tail prevents us from lifting a classical function on this large phase space to a corresponding linear operator in the Hilbert space.

Therefore, the study of classical solutions cannot help in obtaining state-independent mirror operators.

[1] S. D. Mathur, The information paradox: A pedagogical introduction, Classical Quantum Gravity **26**, 224001 (2009).
[2] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, Black holes: complementarity or firewalls?, J. High Energy Phys. 02 (2013) 062.
[3] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, An apologia for firewalls, J. High Energy Phys. 09 (2013) 018.
[4] D. Marolf and J. Polchinski, Gauge-Gravity Duality and the Black Hole Interior, Phys. Rev. Lett. **111**, 171301 (2013).
[5] S. Hawking, Particle creation by black holes, Commun. Math. Phys. **43**, 199 (1975).
[6] S. Hawking, Breakdown of predictability in gravitational collapse, Phys. Rev. D **14**, 2460 (1976).
[7] K. Papadodimas and S. Raju, An infalling observer in AdS/CFT, J. High Energy Phys. 10 (2013) 212.
[8] K. Papadodimas and S. Raju, The Black Hole Interior in AdS/CFT and the Information Paradox, Phys. Rev. Lett. **112**, 051301 (2014).
[9] K. Papadodimas and S. Raju, State-dependent bulk-boundary maps and black hole complementarity, Phys. Rev. D **89**, 086010 (2014).
[10] K. Papadodimas and S. Raju, The unreasonable effectiveness of exponentially suppressed corrections in preserving information, Int. J. Mod. Phys. D **22**, 1342030 (2013).
[11] M. Van Raamsdonk, Evaporating firewalls, J. High Energy Phys. 11 (2014) 038.
[12] D. Harlow, Aspects of the Papadodimas-Raju proposal for the black hole interior, J. High Energy Phys. 11 (2014) 055.
[13] J. Maldacena and L. Susskind, Cool horizons for entangled black holes, Fortschr. Phys. **61**, 781 (2013).
[14] L. Motl, http://motls.blogspot.com/2013/08/one-cant-background-independently.html.
[15] R. Bousso, Observer complementarity upholds the equivalence principle, Phys. Rev. D **87**, 124023 (2013); L. Susskind, The transfer of entanglement: the case for firewalls, arXiv:1210.2098; S. D. Mathur and D. Turton, Comments on black holes I: the possibility of complementarity, J. High Energy Phys. 01 (2014) 034; B. D. Chowdhury and A. Puhm, Is Alice burning or fuzzing? Phys. Rev. D **88**, 063509 (2013); L. Susskind, Singularities, firewalls, and complementarity, arXiv:1208.3445; I. Bena, A. Puhm, and B. Vercnocke, Nonextremal black hole microstates: fuzzballs of fire or fuzzballs of fuzz?, J. High Energy Phys. 12 (2012) 014; A. Giveon and N. Itzhaki, String theory versus black hole complementarity, J. High Energy Phys. 12 (2012) 094; T. Banks and W. Fischler, Holographic spacetime does not predict firewalls, arXiv:1208.4757; A. Ori, Firewall or smooth horizon?, Gen. Relativ. Gravit. **48**, 9 (2016); S. Hossenfelder, Comment on the black hole firewall, arXiv:1210.5317; D.-i. Hwang, B.-H. Lee, and D.-h. Yeom, Is the firewall consistent?: Gedanken experiments on black hole complementarity and firewall proposal, J. Cosmol. Astropart. Phys. 01 (2013) 005; S. G. Avery, B. D. Chowdhury, and A. Puhm, Unitarity and fuzzball complementarity: Alice fuzzes but may not even know it!, J. High Energy Phys. 09 (2013) 012; K. Larjo, D. A. Lowe, and L. Thorlacius, Black holes without firewalls, Phys. Rev. D **87**, 104018 (2013); S. K. Rama, Remarks on black hole evolution a la firewalls and fuzzballs, arXiv:1211.5645; D. N. Page, Hyperentropic gravitational fireballs (grireballs) with firewalls, J. Cosmol. Astropart. Phys. 04 (2013) 037; M. Saravani, N. Afshordi, and R. B. Mann, Empty black holes, firewalls, and the origin of Bekenstein-Hawking entropy, Int. J. Mod. Phys. D **23**, 1443007 (2014); T. Jacobson, Boundary unitarity without firewalls, Int. J. Mod. Phys. D **22**, 1342002 (2013); L. Susskind, Black hole complementarity and the Harlow-Hayden conjecture, arXiv:1301.4505; W. Kim, B.-H. Lee, and D.-h. Yeom, Black hole complementarity and firewall in two dimensions, J. High Energy Phys. 05 (2013) 060; I. Park, On the pattern of black hole information release, Int. J. Mod. Phys. A **29**, 1450047 (2014); S. D. Hsu, Macroscopic superpositions and black hole unitarity, arXiv:1302.0451; S. B. Giddings, Nonviolent information transfer from black holes: a field theory parametrization, Phys. Rev. D **88**, 024018 (2013); B.-H. Lee and D.-h. Yeom, Status report: black hole complementarity controversy, Nucl. Phys. B, Proc. Suppl. **246–247**, 178 (2014); S. G. Avery and B. D. Chowdhury, Firewalls in AdS/CFT, J. High Energy Phys. 10 (2014) 174; B. Kang, Bulk cluster decomposition in AdS/CFT and a no-go theorem for correlators in microstates of extremal black holes, arXiv:1305.2797; B. D. Chowdhury, Black holes vs firewalls and thermofield dynamics, Int. J. Mod. Phys. D **22**, 1342011 (2013); D. N. Page, Excluding black hole firewalls with extreme cosmic censorship, J. Cosmol. Astropart. Phys. 06 (2014) 051; M. Axenides, E. Floratos, and S. Nicolis, Modular discretization of the AdS2/CFT1 holography, J. High Energy Phys. 02 (2014) 109; M. Gary, Still no Rindler firewalls, arXiv:1307.4972; B. D. Chowdhury, Cool horizons lead to information loss, J. High Energy Phys. 10 (2013) 034; A. de la Fuente and R. Sundrum, Holography of the BTZ black hole, inside and out, J. High Energy Phys. 09 (2014) 073; J. L. F. Barbon and E. Rabinovici, Conformal complementarity maps, J. High Energy Phys. 12 (2013) 023; S. Lloyd and J. Preskill, Unitarity of black hole evaporation in final-state projection models, J. High Energy Phys. 08 (2014) 126; S. D. H. Hsu, Factorization of unitarity and black hole firewalls, arXiv:1308.5686; D. N. Page, Time dependence of Hawking radiation entropy, J. Cosmol. Astropart. Phys. 09 (2013) 028; S. B. Giddings and Y. Shi, Effective field theory models for nonviolent information transfer from black holes, Phys. Rev. D **89**, 124032 (2014); S. D. Mathur, What does strong subadditivity tell us about black holes?, Nucl. Phys. B, Proc. Suppl. **251–252**, 16 (2014); S. D. Mathur and D. Turton, The flaw in the firewall argument, Nucl. Phys. **B884**, 566 (2014); V. Balasubramanian, M. Berkooz, S. F. Ross, and J. Simon, Black holes, entanglement, and random matrices, Classical Quantum Gravity **31**, 185009 (2014); B. Freivogel, Energy and information near black hole horizons, J. Cosmol. Astropart. Phys. 07 (2014) 041; R. Akhoury, Unitary S matrices with long-range correlations and the quantum black hole, J. High Energy Phys. 08 (2014) 169; N. Lashkari and J. Simón, From state distinguishability to effective bulk locality, J. High Energy Phys. 06 (2014) 038; J. L. Barbon and E. Rabinovici, Geometry and quantum noise, Fortschr. Phys. **62**, 626 (2014).
[16] E. Verlinde and H. Verlinde, Black hole information as topological qubits, arXiv:1306.0516; Black hole

entanglement and quantum error correction, J. High Energy Phys. 10 (2013) 107.
[17] E. Verlinde and H. Verlinde, Passing through the firewall, arXiv:1306.0515.
[18] E. Verlinde and H. Verlinde, Behind the horizon in AdS/CFT, arXiv:1311.1137.
[19] Y. Nomura, J. Varela, and S. J. Weinberg, Complementarity endures: no firewall for an infalling observer, J. High Energy Phys. 03 (2013) 059; R. Brustein, Origin of the black hole information paradox, Fortschr. Phys. **62**, 255 (2014); R. Brustein and A. Medved, Phases of information release during black hole evaporation, J. High Energy Phys. 02 (2014) 116; Semiclassical black holes expose forbidden charges and censor divergent densities, J. High Energy Phys. 09 (2013) 108; Restoring predictability in semiclassical gravitational collapse, J. High Energy Phys. 09 (2013) 015; E. Silverstein, Backdraft: string creation in an old Schwarzschild black hole, arXiv:1402.1486.
[20] Y. Nomura, J. Varela, and S. J. Weinberg, Black holes, information, and Hilbert space for quantum gravity, Phys. Rev. D **87**, 084050 (2013); Y. Nomura and J. Varela, A note on (no) firewalls: the entropy argument, J. High Energy Phys. 07 (2013) 124; Y. Nomura, J. Varela, and S. J. Weinberg, Black holes or firewalls: a theory of horizons, Phys. Rev. D **88**, 084052 (2013).
[21] S. Braunstein, Black Hole Entropy as Entropy of Entanglement, or its Curtains for the Equivalence principle, Phys. Rev. Lett. **110**, 101301 (2013).
[22] D. Harlow and P. Hayden, Quantum computation vs firewalls, J. High Energy Phys. 06 (2013) 085; L. Susskind, New concepts for old black holes, arXiv:1311.3335.
[23] K. Papadodimas and S. Raju, Local operators in the eternal black hole, arXiv:1502.0669.
[24] S. Ryu and T. Takayanagi, Holographic Derivation of Entanglement Entropy from AdS/CFT, Phys. Rev. Lett. **96**, 181602 (2006); Aspects of holographic entanglement entropy, J. High Energy Phys. 08 (2006) 045.
[25] B. S. DeWitt, Quantum theory of gravity. 1. The canonical theory, Phys. Rev. **160**, 1113 (1967).
[26] H. Bantilan, F. Pretorius, and S. S. Gubser, Simulation of asymptotically AdS5 spacetimes with a generalized harmonic evolution scheme, Phys. Rev. D **85**, 084038 (2012).
[27] V. E. Hubeny, M. Rangamani, and T. Takayanagi, A covariant holographic entanglement entropy proposal, J. High Energy Phys. 07 (2007) 062.
[28] V. Balasubramanian, B. D. Chowdhury, B. Czech, and J. de Boer, Entwinement and the emergence of spacetime, J. High Energy Phys. 01 (2015) 048; V. Balasubramanian, B. D. Chowdhury, B. Czech, J. de Boer, and M. P. Heller, Bulk curves from boundary data in holography, Phys. Rev. D **89**, 086004 (2014).
[29] N. Lashkari, M. B. McDermott, and M. Van Raamsdonk, Gravitational dynamics from entanglement "thermodynamics", J. High Energy Phys. 04 (2014) 195.
[30] T. Faulkner, M. Guica, T. Hartman, R. C. Myers, and M. Van Raamsdonk, Gravitation from entanglement in holographic CFTs, J. High Energy Phys. 03 (2014) 051.
[31] M. Van Raamsdonk, Building up spacetime with quantum entanglement, Gen. Relativ. Gravit. **42**, 2323 (2010); Comments on quantum gravity and entanglement, arXiv:0907.2939.
[32] P. D. Hislop and R. Longo, Modular structure of the local algebras associated with the free massless scalar field theory, Commun. Math. Phys. **84**, 71 (1982); H. Casini, M. Huerta, and R. C. Myers, Towards a derivation of holographic entanglement entropy, J. High Energy Phys. 05 (2011) 036.
[33] M. Nozaki, T. Numasawa, A. Prudenziati, and T. Takayanagi, Dynamics of entanglement entropy from Einstein equation, Phys. Rev. D **88**, 026012 (2013); D. D. Blanco, H. Casini, L.-Y. Hung, and R. C. Myers, Relative entropy and holography, J. High Energy Phys. 08 (2013) 060.
[34] B. Swingle and M. Van Raamsdonk, Universality of gravity from entanglement, arXiv:1405.2933.
[35] D. L. Jafferis and S. J. Suh, The gravity duals of modular Hamiltonians, arXiv:1412.8465.
[36] T. Banks, M. R. Douglas, G. T. Horowitz, and E. J. Martinec, AdS dynamics from conformal field theory, arXiv:hep-th/9808016; I. Bena, On the construction of local fields in the bulk of AdS(5) and other spaces, Phys. Rev. D **62**, 066007 (2000); A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, Holographic representation of local bulk operators, Phys. Rev. D **74**, 066009 (2006); Local bulk operators in AdS/CFT: a boundary view of horizons and locality, Phys. Rev. D **73**, 086003 (2006); Local bulk operators in AdS/CFT and the fate of the BTZ singularity, AMS/IP Stud., Adv. Math. **44**, 85 (2008); D. Kabat, G. Lifschytz, and D. A. Lowe, Constructing local bulk observables in interacting AdS/CFT, Phys. Rev. D **83**, 106009 (2011).
[37] R. Bousso, B. Freivogel, S. Leichenauer, V. Rosenhaus, and C. Zukowski, Null geodesics, local CFT operators and AdS/CFT for subregions, Phys. Rev. D **88**, 064057 (2013); S.-J. Rey and V. Rosenhaus, Scanning tunneling macroscopy, black holes, and AdS/CFT bulk Locality, J. High Energy Phys. 07 (2014) 050.
[38] I. A. Morrison, Boundary to bulk maps for AdS causal wedges and the Reeh-Schlieder property in holography, J. High Energy Phys. 05 (2014) 053.
[39] S. Banerjee, K. Papadodimas, S. Raju, and P. Samantray (to be published).
[40] H. Lin, Almost Commuting Selfadjoint Matrices and Applications, *Operator Algebras and Their Applications*, edited by P. A. Fillmore and J. A. Mingo (Fields Institute for Research in Mathematical Sciences, Waterloo, ON, 1997), Vol. 13, p. 193; P. Friis and M. Rørdam, Almost commuting self-adjoint matrices-a short proof of huaxin lin's theorem, J. Reine Angew. Math. **479**, 121 (1996).
[41] A. C. Wall, A proof of the generalized second law for rapidly evolving Rindler horizons, Phys. Rev. D **82**, 124019 (2010); R. Bousso, H. Casini, Z. Fisher, and J. Maldacena, Proof of a quantum Bousso bound, Phys. Rev. D **90**, 044002 (2014).
[42] M. Guica and S. F. Ross, Behind the geon horizon, Classical Quantum Gravity **32**, 055014 (2015).
[43] J. M. Deutsch, Quantum statistical mechanics in a closed system, Phys. Rev. A **43**, 2046 (1991); M. Srednicki, The approach to thermal equilibrium in quantized chaotic

systems, J. Phys. A **32**, 1163 (1999); Chaos and quantum thermalization, Phys. Rev. E **50**, 888 (1994).
[44] H. Lin, Almost commuting self-adjoint matrices and applications, operator algebras and their applications, Waterloo, ON, 1994/1995 **13**, 193 (1997); P. Friis and M. Rørdam, Almost commuting self-adjoint matrices—a short proof of huaxin lin's theorem, J. Reine Angew. Math. 121 (1996).
[45] G. 't Hooft, On the quantum structure of a black hole, Nucl. Phys. **B256**, 727 (1985).
[46] S. Raju, in the Talk at the KITP Workshop on Fuzz, Fire, or Complementarity, 2013.
[47] D. Marolf and A. C. Wall, Eternal black holes and superselection in AdS/CFT, Classical Quantum Gravity **30**, 025001 (2013).
[48] T. Andrade, S. Fischetti, D. Marolf, S. F. Ross, and M. Rozali, Entanglement and correlations near extremality: CFTs dual to Reissner-Nordström AdS_5, J. High Energy Phys. 04 (2014) 023.
[49] T. Regge and C. Teitelboim, Role of surface integrals in the Hamiltonian formulation of general relativity, Ann. Phys. (N.Y.) **88**, 286 (1974).
[50] J. D. Brown and M. Henneaux, Central charges in the canonical realization of asymptotic symmetries: an example from three-dimensional gravity, Commun. Math. Phys. **104**, 207 (1986).
[51] M. Guica, T. Hartman, W. Song, and A. Strominger, The Kerr/CFT correspondence, Phys. Rev. D **80**, 124008 (2009).
[52] S. G. Avery and B. D. Chowdhury, No holography for eternal AdS black holes, arXiv:1312.3346.
[53] S. D. Mathur, What is the dual of two entangled CFTs?, arXiv:1402.6378.
[54] L. Susskind, L. Thorlacius, and J. Uglum, The stretched horizon and black hole complementarity, Phys. Rev. D **48**, 3743 (1993).
[55] S. Raju, in Talk at Strings 2014, Princeton.
[56] S. Ghosh and S. Raju (to be published).
[57] S. Banerjee, J.-W. Bryan, K. Papadodimas, and S. Raju, A toy model of black hole complementarity (to be published).
[58] J. Polchinski, in Talk at Strings 2014, Princeton.
[59] E. T. Jaynes, Information theory and statistical mechanics, Phys. Rev. **106**, 620 (1957); Information theory and statistical mechanics. ii, Phys. Rev. **108**, 171 (1957).
[60] D. Marolf and J. Polchinski, Violations of the Born rule in cool state-dependent horizons, Phys. Rev. D **48**, 3743 (1993).
[61] S. H. Shenker and D. Stanford, Multiple shocks, J. High Energy Phys. 12 (2014) 046.
[62] D. Bak, M. Gutperle, and R. A. Janik, Janus black holes, J. High Energy Phys. 10 (2011) 056; D. Bak, M. Gutperle, and A. Karch, Time-dependent black holes and thermal equilibration, J. High Energy Phys. 12 (2007) 034; D. Bak, M. Gutperle, and S. Hirano, Three-dimensional Janus and time-dependent black holes, J. High Energy Phys. 02 (2007) 068.
[63] N. Gisin, Weinberg's nonlinear quantum mechanics and supraluminal communications, Phys. Lett. **143A**, 1 (1990).
[64] J. Polchinski, Weinberg's Nonlinear Quantum Mechanics and the Einstein-Podolsky-Rosen Paradox, Phys. Rev. Lett. **66**, 397 (1991).
[65] S. Weinberg, Testing quantum mechanics, Ann. Phys. (N.Y.) **194**, 336 (1989).
[66] S. H. Shenker and D. Stanford, Black holes and the butterfly effect, J. High Energy Phys. 03 (2014) 067; J. Maldacena, S. H. Shenker, and D. Stanford, A bound on chaos, arXiv:1503.0140.
[67] L. G. Yaffe, Large n limits as classical mechanics, Rev. Mod. Phys. **54**, 407 (1982).
[68] E. Sudarshan, Equivalence of Semiclassical and Quantum Mechanical Descriptions of Statistical Light Beams, Phys. Rev. Lett. **10**, 277 (1963); C. Mehta, Diagonal Coherent-State Representation of Quantum Operators, Phys. Rev. Lett. **18**, 752 (1967).
[69] P. Dedecker, in *Geometrie différentielle, Colloq. Intern. du CNRS LII, Strasbourg* (1953), p. 17; H. Goldschmidt and S. Sternberg, The Hamilton-Cartan formalism in the calculus of variations, Ann. Inst. Fourier **23**, 203 (1973); J. Kijowski, A finite-dimensional canonical formalism in the classical field theory, Commun. Math. Phys. **30**, 99 (1973); K. Gawedzki and W. Kondracki, Canonical formalism for the local-type functionals in the classical field theory, Rep. Math. Phys. **6**, 465 (1974); W. Szczyrba, A symplectic structure on the set of Einstein metrics, Commun. Math. Phys. **51**, 163 (1976); P. Garcia, Reducibility of the symplectic structure of classical fields with gauge symmetry, Lect. Notes Math. **570**, 365 (1977); G. Zuckerman, Action principles and global geometry, *Mathematical Aspects of String Theory* (World Scientific, Singapore, 1987), Vol. 1, p. 259.
[70] C. Crnkovic and E. Witten, Covariant Description of Canonical formalism in Geometrical Theories, *Three Hundred Years of Gravitation* (Cambridge University Press, Camridge, 1987), pp. 676–684.
[71] J. Lee and R. M. Wald, Local symmetries and constraints, J. Math. Phys. **31**, 725 (1990).
[72] D. N. Page and K. Phillips, Self-gravitating radiation in anti–de Sitter space, Gen. Relativ. Gravit. **17**, 1029 (1985).